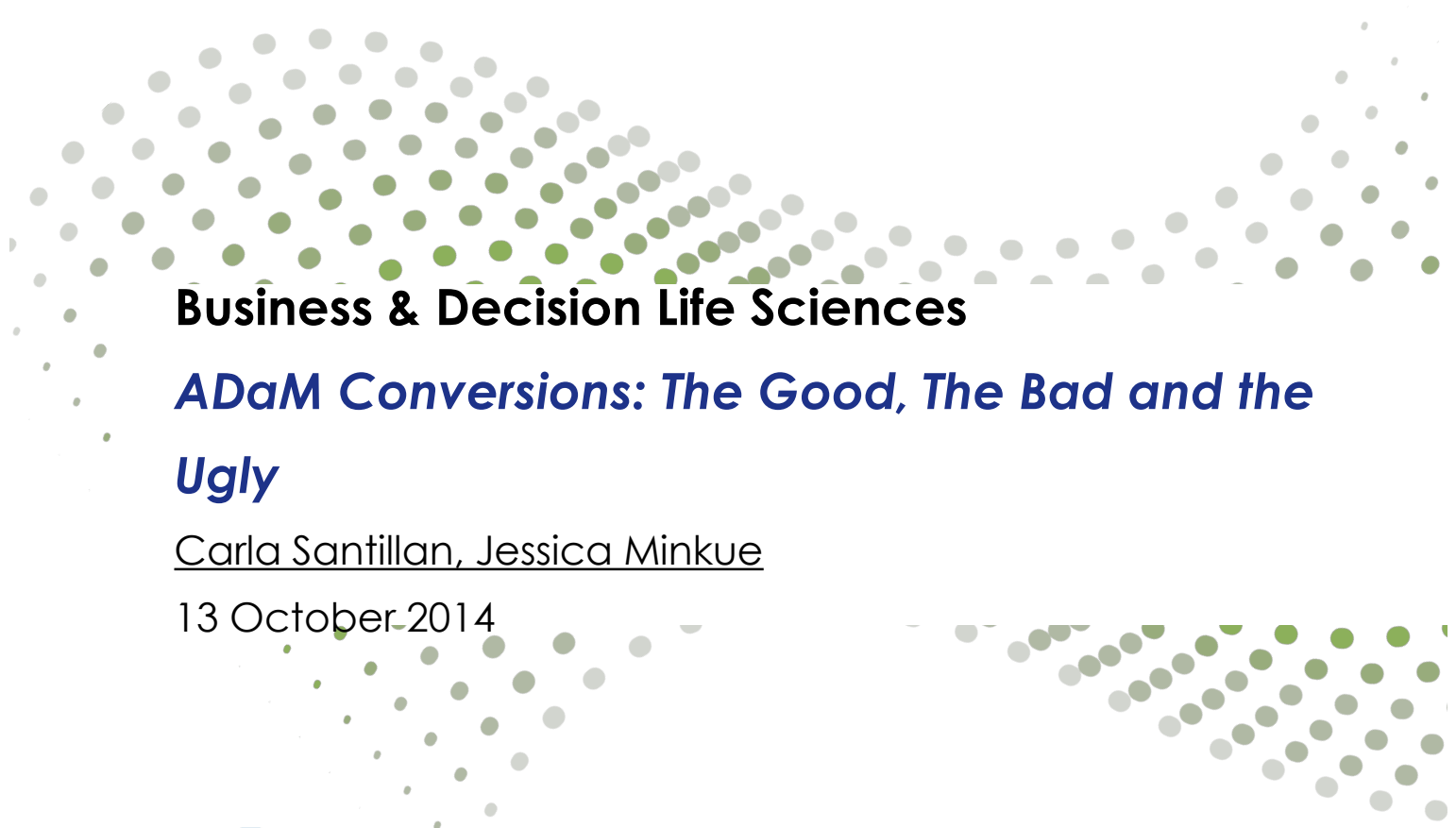




Business & Decision Life Sciences

ADaM Conversions: The Good, The Bad and the Ugly

Carla Santillan, Jessica Minkue

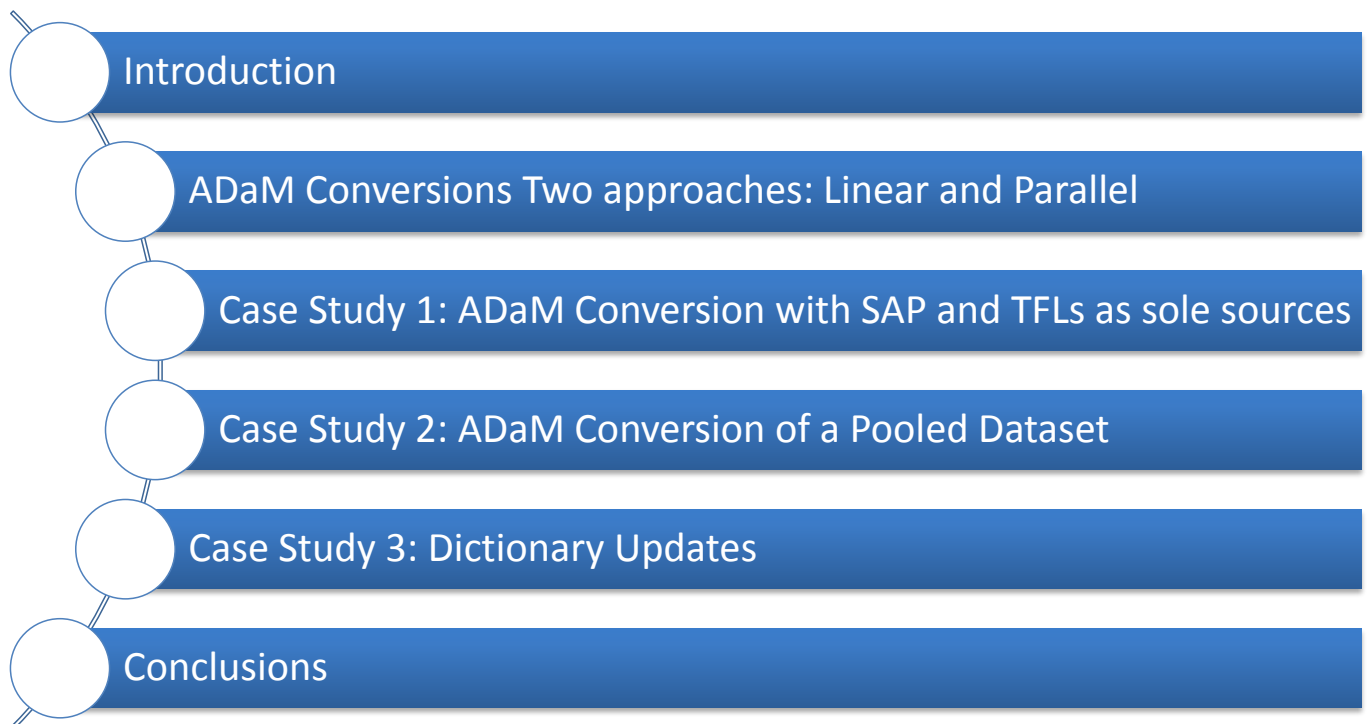
13 October 2014



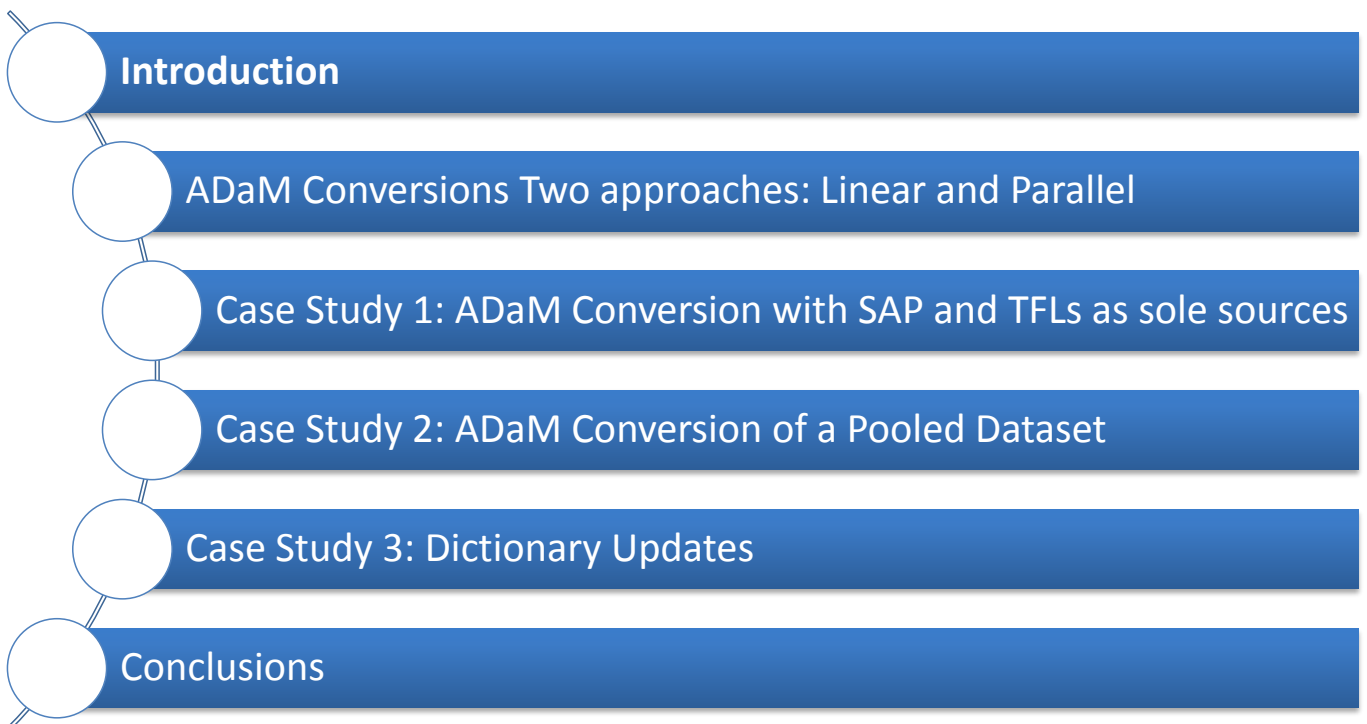
Restricted © Business & Decision Life Sciences 2014 All rights reserved.



ADaM Conversions: The Good, the Bad and the Ugly



ADaM Conversions: The Good, the Bad and the Ugly



Introduction



Business & Decision Life Sciences' experiences in ADaM conversion projects applying:

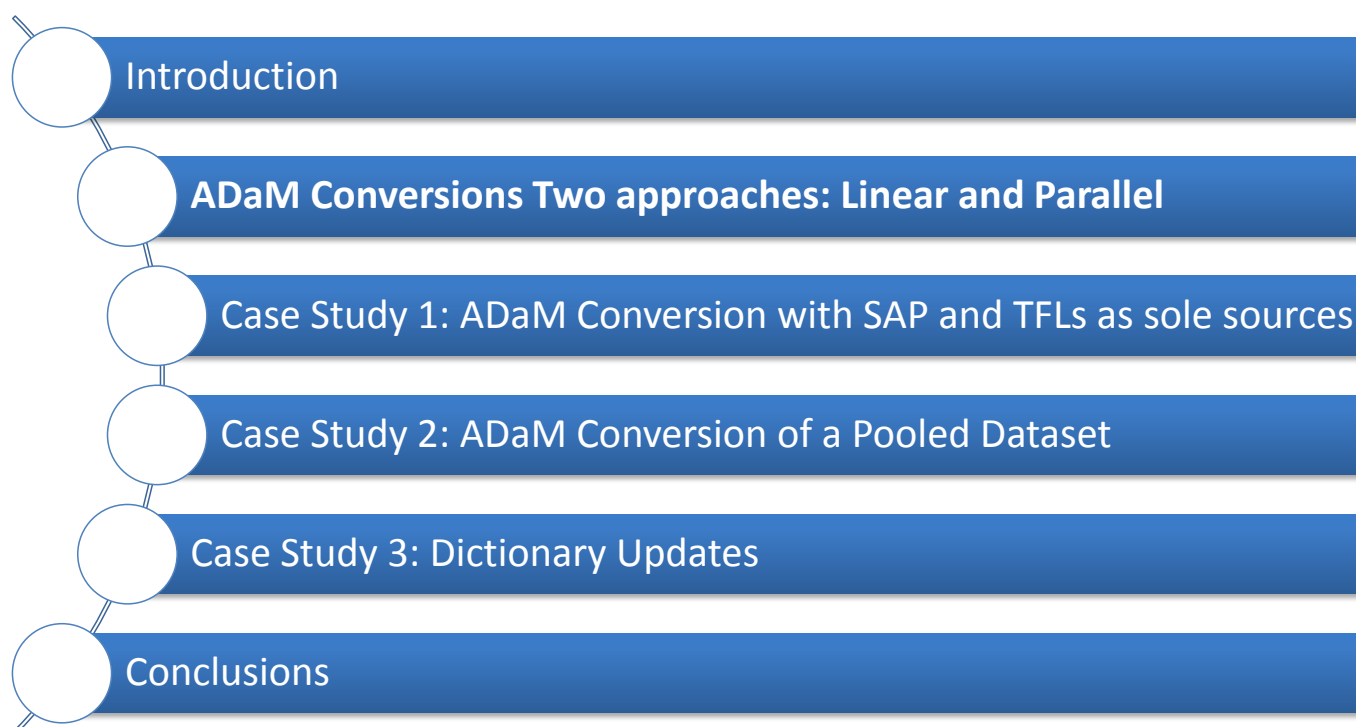
- Linear approach
- Parallel approach for Pooled datasets
- Dictionary updates in a Pooled database

Objective

Describing the good, bad and ugly experiences gained during the execution of these projects



ADaM Conversions: The Good, the Bad and the Ugly



ADaM Conversions Two approaches: Linear and Parallel



Good understanding of the relationship between

- Analysis results
- Analysis datasets
- SDTM domains

The **core** of any ADaM conversion is the SDTM/ADaM **traceability**, this can be found at two levels:

- **Metadata traceability:** finding a relationship between an analysis result and analysis dataset(s), or a relationship of the analysis variable to its source dataset(s) and variable(s)
- **Data point traceability:** finding the predecessor record(s)



ADaM Conversions Two approaches: Linear and Parallel



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	28.49	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%)

• Patient Demographics - Part 1

Patient X5 Page 4 (Demo: Via for Visit 1); Page 4 of 4

Visit Date: 11-Dec-2007

Blank Comment

PATIENT DEMOGRAPHICS - Part 1

Inform consent was obtained on: **DSSTDC**

Consent: **SEX** 1 = male, 2 = female

Date of birth: **BRTHDT** Age: **AGE** years **AGEU**

(Age is automatically calculated when screen is saved and closed)

Height: cm

Weight: kg

Waist circumference: cm

V\$ORRES / V\$ORRESU where V\$TESTCD = "HEIGHT", "WEIGHT", "WAIST"

	STUDYID	USUBID	SUBID	BMI	BMI01	BMI02	BMI03	BMI04
2	9999-0001	9999-0001-000001	000001	27.77777778	<30 kg/m**2	1.25<-30 kg/m**2	2	2
3	9999-0001	9999-0001-000002	000002	25.02615762	<30 kg/m**2	1.25<-30 kg/m**2	2	2
4	9999-0001	9999-0001-000003	000003	26.17618421	<30 kg/m**2	1.25<-30 kg/m**2	2	2
5	9999-0001	9999-0001-000004	000004	36.192625	>=30 kg/m**2	2. >=30 kg/m**2	3	3
6	9999-0001	9999-0001-000005	000005	30.86885931	>=30 kg/m**2	2. >=30 kg/m**2	3	3
7	9999-0001	9999-0001-000006	000006	38.89716291	>=30 kg/m**2	2. >=30 kg/m**2	3	3
8	9999-0001	9999-0001-000007	000007	25.93446261	<30 kg/m**2	1.25<-30 kg/m**2	2	2
9	9999-0001	9999-0001-000008	000008	30.10386228	>=30 kg/m**2	2. >=30 kg/m**2	3	3
10	9999-0001	9999-0001-000009	000009	32.30266261	>=30 kg/m**2	2. >=30 kg/m**2	3	3
11	9999-0001	9999-0001-000010	000010	28.97613787	<30 kg/m**2	1.25<-30 kg/m**2	2	2
12	9999-0001	9999-0001-000011	000011	29.372297363	<30 kg/m**2	1.25<-30 kg/m**2	2	2
13	9999-0001	9999-0001-000012	000012	26.74465268	<30 kg/m**2	1.25<-30 kg/m**2	2	2
14	9999-0001	9999-0001-000013	000013	32.71861863	>=30 kg/m**2	2. >=30 kg/m**2	3	3
15	9999-0001	9999-0001-000014	000014	28.71972183	<30 kg/m**2	1.25<-30 kg/m**2	2	2
16	9999-0001	9999-0001-000015	000015	32.27042577	>=30 kg/m**2	2. >=30 kg/m**2	3	3

Variable	Label	Type	Controlled Terminology	Origin	Role	Comment
STUDYID	Study Identifier	text	Protocol, CRF Page 1	Assigned	IDENTIFIER	The STUDYID variable has a fixed format 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code
USUBID	Unique Subject Identifier	text	Protocol	Assigned	IDENTIFIER	The USUBID variable has a fixed format 'XXXX-YYYY-ZZZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZ' the 5-digit patient code
V\$SEQ	Sequence Number	integer	Derived	Derived	IDENTIFIER	Sequence number (automatically generated to ensure uniqueness within a dataset for a subject)
V\$TESTCD	Visit Sign Test Short Name	text	Assigned	Derived	TOPE	
V\$TESTC	Visit Sign Test Name	text	Derived	Derived	ECONOMY QUALIFIER	
V\$POS	Visit Sign Position of Subject	text	Derived, CRF Page 1, 11	Derived	RESULT QUALIFIER	
V\$ORRES	Result or Finding in Original Data	text	Derived, CRF Page 1, 11	Derived	RESULT QUALIFIER	
V\$ORRESU	Original Data	text	Derived, CRF Page 1, 11	Derived	RESULT QUALIFIER	
V\$TESTC	Character Result Finding in Test Format	text	Derived	Derived	RESULT QUALIFIER	
V\$TESTSU	Character Result Finding in Standard Data	text	Derived	Derived	RESULT QUALIFIER	

Variable	Label	Type	Controlled Terminology	Controlled Terminology	Origin	Role	Comment
STUDYID	Study Identifier	text	Protocol, CRF Page 1	Assigned	IDENTIFIER	IDENTIFIER	The STUDYID variable has a fixed format 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code
USUBID	Unique Subject Identifier	text	Protocol	Assigned	IDENTIFIER	IDENTIFIER	The USUBID variable has a fixed format 'XXXX-YYYY-ZZZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZ' the 5-digit patient code
HEIGHT	Subject Height (cm)	integer	Derived	Derived	ANALYSIS	ANALYSIS	
WEIGHT	Subject Weight (kg)	integer	Derived	Derived	ANALYSIS	ANALYSIS	
BMI	Baseline BMI (kg/m**2)	integer	Derived	Derived	ANALYSIS	ANALYSIS	
BMI01	Category 1 of Baseline BMI	text	Derived	Derived	ANALYSIS	ANALYSIS	
BMI02	Category 2 of Baseline BMI	text	Derived	Derived	ANALYSIS	ANALYSIS	
BMI03	Category 3 of Baseline BMI	text	Derived	Derived	ANALYSIS	ANALYSIS	
BMI04	Category 4 of Baseline BMI	text	Derived	Derived	ANALYSIS	ANALYSIS	
BMI05	Category 5 of Baseline BMI	text	Derived	Derived	ANALYSIS	ANALYSIS	

Business & Decision
Life Sciences

	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%)

Patient Demographics – Part I

Patent 302 Page 4 (Home Visit Visit 50) Page 1 of 1.

Visit Date **11-Dec-2007** Email Comment

PATIENT DEMOGRAPHICS – Part I **DSCAT = "PROTOCOL MILESTONE"**

Information contained was obtained on **DSSTDTC**

DATE1 = **DSDATE** **DSOEDATE**

Gender: **SEX** 1 = male, 2 = female

Date of birth: **BIRTHDT** Age years A

(Age is automatically calculated when born saved and closed)

Height cm

Weight kg

Waist circumference cm

V\$ORRES1" V\$ORRESU where
V\$TESTED" = "HEIGHT,"
"WEIGHT," "WAIST"

Vital Signs Database (VS)						Table: vital_signs
Variable	Label	Type	Controlled Terminology	Origin	Role	Comments
STUDID	Study Identifier	text	Protocol, CRF Page 1		IDENTIFIER	The STUDID variable has a fixed format: NNNN.YYYY, where NNNN indicates the 4-digit compound code and the YYYY is the 4-digit study code.
DOMAIN	Disease Alteration	text	DOMAIN	Assigned	IDENTIFIER	
USUBID	Subject Identifier	text	Protocol		IDENTIFIER	The USUBID variable has a fixed format: NNNN.YYYY.ZZZZZZ, where NNNN is the 4-digit study code and ZZZZZZ is the 6-digit patient code.
VSSEQ	Sequence Number	integer	Derived		IDENTIFIER	Sequence number (automatically generated) to ensure sequences within a dataset for a subject
VSTEST	Vital Signs Test Short Name	text	Assigned		TOPIC	
VSTEST	Vital Signs Test Name	text	Derived		32000M QUALIFIER	
VSPON	Vital Signs Position of Subject	text	CRF Page 1		RECORD QUALIFIER	
VORRES	Result or Finding in Original Units	text	Derived, CRF Page 5, 11		RESULT QUALIFIER	
VORRESUC	Original Units	text	VARIABLE	CRF Page 5, 11	VARIABLE QUALIFIER	
VSTESTSC	Character String Finding in Unit Format	text	Derived		RESULT QUALIFIER	
VSTESTSC	Numeric Result Finding	float	3, 1	Derived	RESULT	

Subject Level Analysis Dataset Dataset (SMLD)							..AnalyticsHub
Variable	Label	Type	Composited Form of	Computation Algorithm or Method	Origin	Role	Comment
STUDID	Study Identifier	text			DML	IDENTIFIER	The STUDID variable has a four-letter 'XX' YYYY, where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit patient code.
STUDID2	Unique Subject Identifier	text			DML	IDENTIFIER	The STUDID2 variable has a four-letter 'XXX' YYYY-ZZZZZ, where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit code, and 'ZZZZZ' the 5-digit patient code.
SUBID	Subject Identifier for the Study	text			DML	IDENTIFIER	
HEIGHT	Baseline Height (cm)	integer		ANAL_HEIGHT	Derived	ANAL_VSN	
WEIGHT	Baseline Weight (kg)	integer		ANAL_WEIGHT	Derived	ANAL_VSN	
BMI	Baseline BMI (kg/m**2)	integer		ANAL_BMI	Derived	ANAL_VSN	
BMGR1	Category 1 of Baseline BMI	integer	BMGR1_L	ANAL_BMGR1	Derived	ANAL_VSN	
BMGR1N	Category 1 of Baseline BMGR1N	integer	BMGR1N_L	ANAL_BMGR1N	Derived	ANAL_VSN	
BMGR2	Category 2 of Baseline BMI	integer	BMGR2_L	ANAL_BMGR2	Derived	ANAL_VSN	
BMGR2N	Category 2 of Baseline BMGR2N	integer	BMGR2N_L	ANAL_BMGR2N	Derived	ANAL_VSN	
BMGR3	Category 3 of Baseline BMI	integer	BMGR3_L	ANAL_BMGR3	Derived	ANAL_VSN	
BMGR3N	Category 3 of Baseline BMGR3N	integer	BMGR3N_L	ANAL_BMGR3N	Derived	ANAL_VSN	

ADaM Conversions Two approaches: Linear and Parallel



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	28.49	28.47
Max	40.1	41.2
Baseline BMI (categorical) [N (%)]		
<25 kg/m**2	41 (24.6%)	71 (21.1%)

• Patient Demographics - Part 1

Visit Date 11-Dec-2007

PATIENT DEMOGRAPHICS - Part 1

Inform consent was obtained on **DSSTDC**

Gender **SEX** **DSSTDC**

Date of birth **BRTHDC** Age **AGE** years **AGEU**

1 = male, 2 = female
(Age is automatically calculated when screen is saved and closed)

Height **HT** cm

Weight **WT** kg

Waist circumference **WC** cm

VSTESTCD = "HEIGHT", "WEIGHT", "WAIST"

	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	28.876133787	<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	32.718619869	>=30 kg/m**2	2	>=30 kg/m**2	3
15	9999-0001	9999-0001-000014	000014	32.719723183	<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

Variable	Label	Type	Conformed Terminology	Domain	Source	Notes
STUDYID	Study Identifier	text	Protocol CRF Page 1	IDENTIFIER	DM	The STUDYID variable has a fixed format 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code.
USUBJID	Unique Subject Identifier	text	Protocol	IDENTIFIER	DM	The USUBJID variable has a fixed format 'XXXX-YYYY-ZZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZ' the 5-digit patient code.
VSSEQ	Sequence Number	integer	Desired	IDENTIFIER	DM	Sequence number (automatically generated to ensure uniqueness within a dataset for a subject).
VSTESTCD	Visit Sign Test Short Name	text	Assigned	TEST	DM	
VSTEST	Visit Sign Test Name	text	Desired	TEST	DM	
VSPON	Visit Sign Position of Subject	text	Desired, CRF Page 1, 11	TEST	DM	
VSPONRES	Result or Finding in Original Data	text	Desired, CRF Page 1, 11	TEST	DM	
VSPONRES2	Original Data	text	Desired, CRF Page 1, 11	TEST	DM	
VSTESTCD	Character Format Flagging in Test Format	text	Desired	TEST	DM	
VSTESTRES	Character Format Flagging in Standard Data	text	Desired	TEST	DM	

Variable	Label	Type	Conformed Terminology	Domain	Source	Notes
STUDYID	Study Identifier	text	Protocol CRF Page 1	IDENTIFIER	DM	The STUDYID variable has a fixed format 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code.
USUBJID	Unique Subject Identifier	text	Protocol	IDENTIFIER	DM	The USUBJID variable has a fixed format 'XXXX-YYYY-ZZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZ' the 5-digit patient code.
VSSEQ	Sequence Number	integer	Desired	IDENTIFIER	DM	Sequence number (automatically generated to ensure uniqueness within a dataset for a subject).
VSSEQ	Sequence Number	integer	Desired	IDENTIFIER	DM	Sequence number (automatically generated to ensure uniqueness within a dataset for a subject).
VSSEQ	Sequence Number	integer	Desired	IDENTIFIER	DM	Sequence number (automatically generated to ensure uniqueness within a dataset for a subject).
VSSEQ	Sequence Number	integer	Desired	IDENTIFIER	DM	Sequence number (automatically generated to ensure uniqueness within a dataset for a subject).
VSSEQ	Sequence Number	integer	Desired	IDENTIFIER	DM	Sequence number (automatically generated to ensure uniqueness within a dataset for a subject).
VSSEQ	Sequence Number	integer	Desired	IDENTIFIER	DM	Sequence number (automatically generated to ensure uniqueness within a dataset for a subject).
VSSEQ	Sequence Number	integer	Desired	IDENTIFIER	DM	Sequence number (automatically generated to ensure uniqueness within a dataset for a subject).
VSSEQ	Sequence Number	integer	Desired	IDENTIFIER	DM	Sequence number (automatically generated to ensure uniqueness within a dataset for a subject).

Business & Decision
Life Sciences

ADaM Conversions Two approaches: Linear and Parallel



Table 1 Demographic Data - Per-Protocol

	Treatment 1	Treatment 2
--	-------------	-------------

Baseline body mass index (BMI) [kg/m**2]

Subject-Level Analysis Dataset Dataset (ADSL)							Analysis/adsl.xpt
Variable	Label	Type	Controlled Terms or Format	Computational Algorithm or Method	Origin	Role	Comment
STUDYID	Study Identifier	text			DM	IDENTIFIER	The STUDYID variable has a fixed format 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code
USUBJID	Unique Subject Identifier	text			DM	IDENTIFIER	The USUBJID variable has a fixed format 'XXXX-YYYY-ZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZ' the 5-digit patient code
SUBJID	Subject Identifier for the Study	text			DM	IDENTIFIER	
HEIGHT	Baseline Height (cm)	integer		ADSL HEIGHT	Derived	ANALYSIS	
WEIGHT	Baseline Weight (kg)	integer		ADSL WEIGHT	Derived	ANALYSIS	
BMI	Baseline BMI (kg/m**2)	integer		ADSL BMI	Derived	ANALYSIS	
BMIGR1	Category 1 of Baseline BMI	text	BMIGR1L	ADSL BMIGR1	Derived	ANALYSIS	
BMIGR1N	Category 1 of Baseline BMI, (N)	integer	BMIGR1N	ADSL BMIGR1N	Derived	ANALYSIS	
BMIGR2	Category 2 of Baseline BMI	text	BMIGR2L	ADSL BMIGR2	Derived	ANALYSIS	
BMIGR2N	Category 2 of Baseline BMI, (N)	integer	BMIGR2N	ADSL BMIGR2N	Derived	ANALYSIS	
BMIGR3	Category 3 of Baseline BMI	text	BMIGR3L	ADSL BMIGR3	Derived	ANALYSIS	
BMIGR3N	Category 3 of Baseline BMI, (N)	integer	BMIGR3N	ADSL BMIGR3N	Derived	ANALYSIS	

• Patient Demographics - Part 1
Patient X5 Page 4 (Demo: Via for Visit 1): Page 4 of 4
Visit Date [11-Dec-2007] Blank Comment

PATIENT DEMOGRAPHICS - Part 1
Informed consent was obtained on [DSCTERM / DSDECD] [DSSTDTG]
Gender [SEX] 1 = male, 2 = female
Date of birth [BRTHDTC] Age [AGE] years
(Age is automatically calculated when saved and closed)
Height [] cm
Weight [] kg
Waist circumference [] cm
VSORRES / VSORRESU where
VSTESTCD = "HEIGHT",
"WEIGHT", "WAIST"

Variable	Label	Type	Controlled Terminology	Part
STUDYID	Study Identifier	text		Part
DOMAIN	Domain Identification	text	DOMAIN	Part
USUBJID	Unique Subject Identifier	text		Part
VISSEQ	Sequence Number	integer		Part
VISITNUM	Visit Sign Test Short Name	text		Part
VISITCD	Visit Sign Test Name	text	Derived	Part
VISITPOS	Visit Sign Position of Subject	text	POSITION	Part
VISITRES	Result or Finding in Original Data	text	Derived, CRF Page 1, 11	Part
VISITRES2	Original Data	text	UNRESL	Part
VISITRES3	Character Fetched in Full Format	text	Derived	Part
VISITRES4	Character Fetched in Standard Data	text	Derived	Part

Variable	Label	Type	Controlled Terminology	Part
BMIGR1	Category 1 of Baseline BMI	text	BMIGR1L	Part
BMIGR1N	Category 1 of Baseline BMI, (N)	integer	BMIGR1N	Part
BMIGR2	Category 2 of Baseline BMI	text	BMIGR2L	Part
BMIGR2N	Category 2 of Baseline BMI, (N)	integer	BMIGR2N	Part
BMIGR3	Category 3 of Baseline BMI	text	BMIGR3L	Part
BMIGR3N	Category 3 of Baseline BMI, (N)	integer	BMIGR3N	Part

ADaM Conversions Two approaches: Linear and Parallel



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	28.49	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%)

• Patient Demographics - Part I

Patient X5 Page 4 (Demo, Via for Visit 1); Page 4 of 4.

Visit Date: 11-Dec-2007

Vital Signs Dataset (VS)

PATIENT DEMOGRAPHICS - Part I
Informed consent was obtained
OS TERM / OS DECOD
Gender: SEX
Date of birth: BIRTHDT
Height
Weight
Waist circumference
VSORRES / VSORRES
VSTESTCD = "HEIGHT"
"WEIGHT", "WAIST"

Variable	Label	Type	Controlled Terminology	Origin	Role	Comment
STUDYID	Study Identifier	text		Protocol, CRF Page 1	IDENTIFIER	The STUDYID variable has a fixed format: 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code
DOMAIN	Domain Abbreviation	text	DOMAIN	Assigned	IDENTIFIER	
USUBJID	Unique Subject Identifier	text		Protocol	IDENTIFIER	The USUBJID variable has a fixed format: 'XXXX-YYYY-ZZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZ' the 6-digit patient code
VSSEQ	Sequence Number	integer		Derived	IDENTIFIER	Sequence number (automatically generated) to ensure uniqueness within a dataset for a subject
VSTESTCD	Vital Signs Test Short Name	text		Assigned	TOPIC	
VSTEST	Vital Signs Test Name	text		Derived	SYNONYM QUALIFIER	
VSPOS	Vital Signs Position of Subject	text	POSITION	CRF Page 13	RECORD QUALIFIER	
VSORRES	Result or Finding in Original Units	text		Derived, CRF Page 9, 13	RESULT QUALIFIER	
VSORRESU	Original Units	text	VSRESU	CRF Page 9, 13	VARIABLE QUALIFIER	
VSTRESC	Character Result Finding in Std Format	text		Derived	RESULT QUALIFIER	
VSTRESN	Numeric Result Finding in Standard Units	float	3.1	Derived	RESULT QUALIFIER	



..Tabulations/vs.xpt

BMISR1	BMISR1N	BMISR2	BMISR2N
7778 (<30 kg/m**2)	1	25-<30 kg/m**2	2
15102 (<30 kg/m**2)	1	25-<30 kg/m**2	2
14521 (<30 kg/m**2)	1	25-<30 kg/m**2	2
19626 (>=30 kg/m**2)	2	>=30 kg/m**2	3
56131 (>=30 kg/m**2)	2	>=30 kg/m**2	3
65916 (>=30 kg/m**2)	2	>=30 kg/m**2	3
46281 (<30 kg/m**2)	1	25-<30 kg/m**2	2
46228 (>=30 kg/m**2)	2	>=30 kg/m**2	3
62693 (>=30 kg/m**2)	2	>=30 kg/m**2	3
13787 (<30 kg/m**2)	1	25-<30 kg/m**2	2
97383 (<30 kg/m**2)	1	25-<30 kg/m**2	2
65598 (<30 kg/m**2)	1	25-<30 kg/m**2	2
19893 (>=30 kg/m**2)	2	>=30 kg/m**2	3
23183 (<30 kg/m**2)	1	25-<30 kg/m**2	2
20277 (>=30 kg/m**2)	2	>=30 kg/m**2	3

..Analyze/vs.xpt

Comment

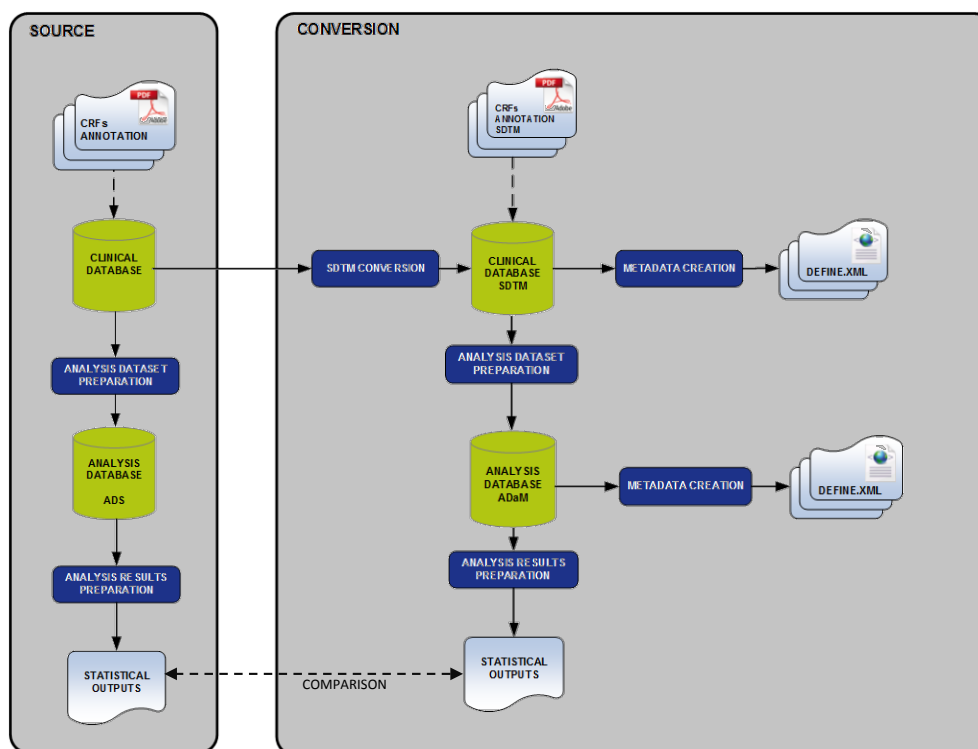
Variable has a fixed format 'XXXX-YYYY-ZZZZZ'.

'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code.

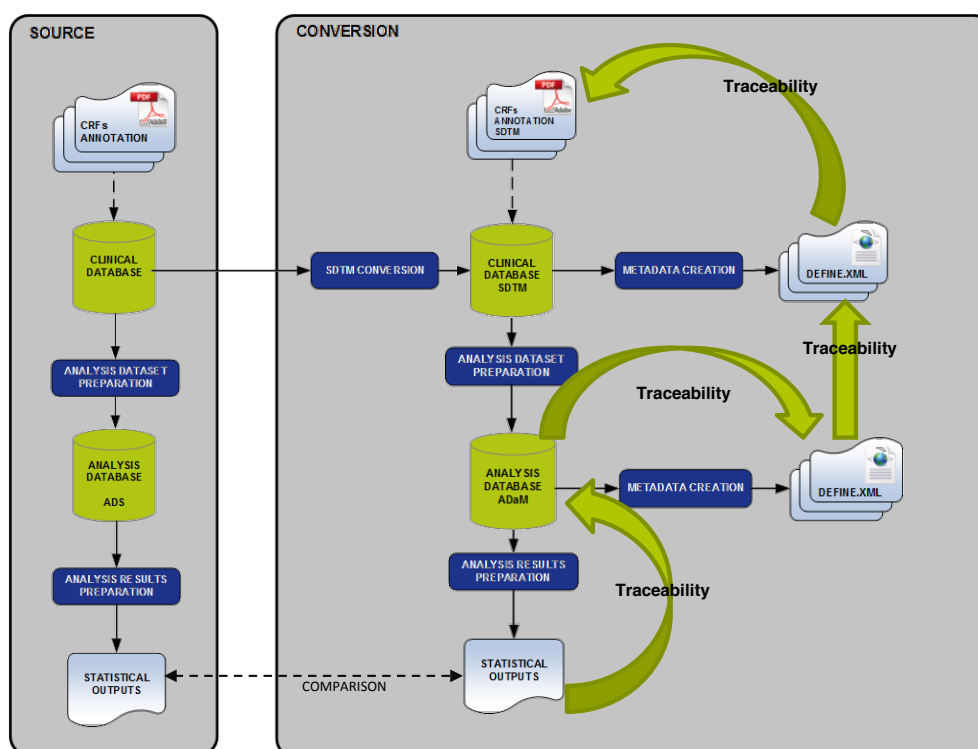
Variable has a fixed format 'XXXX-YYYY-ZZZZZ'.

'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZ' the 6-digit patient code.

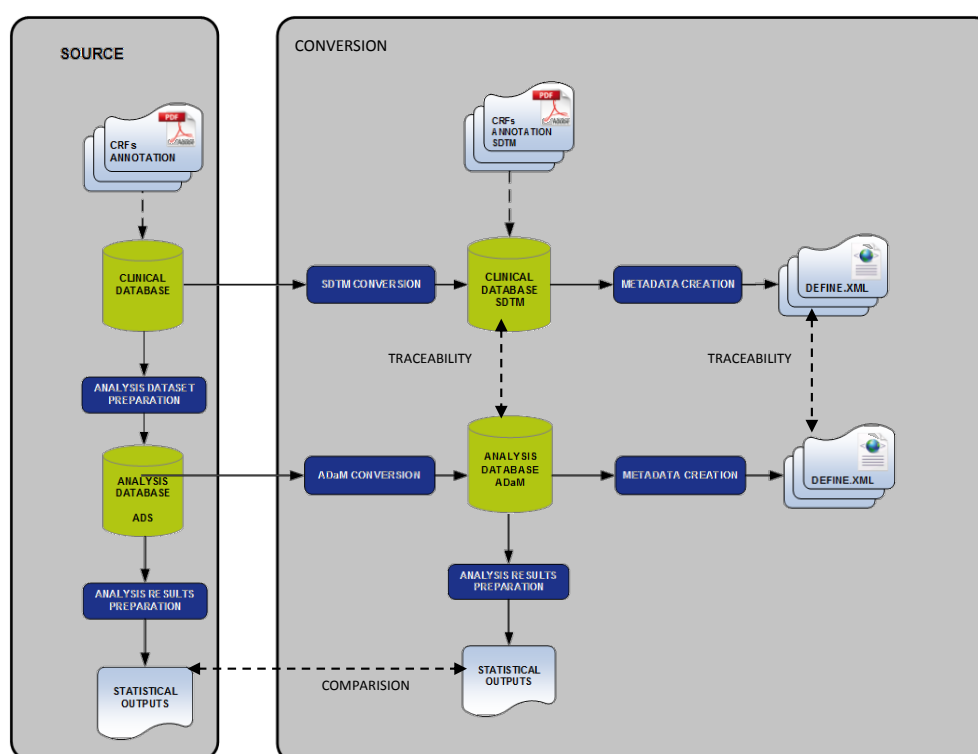
ADaM Conversions Two approaches: Linear and Parallel



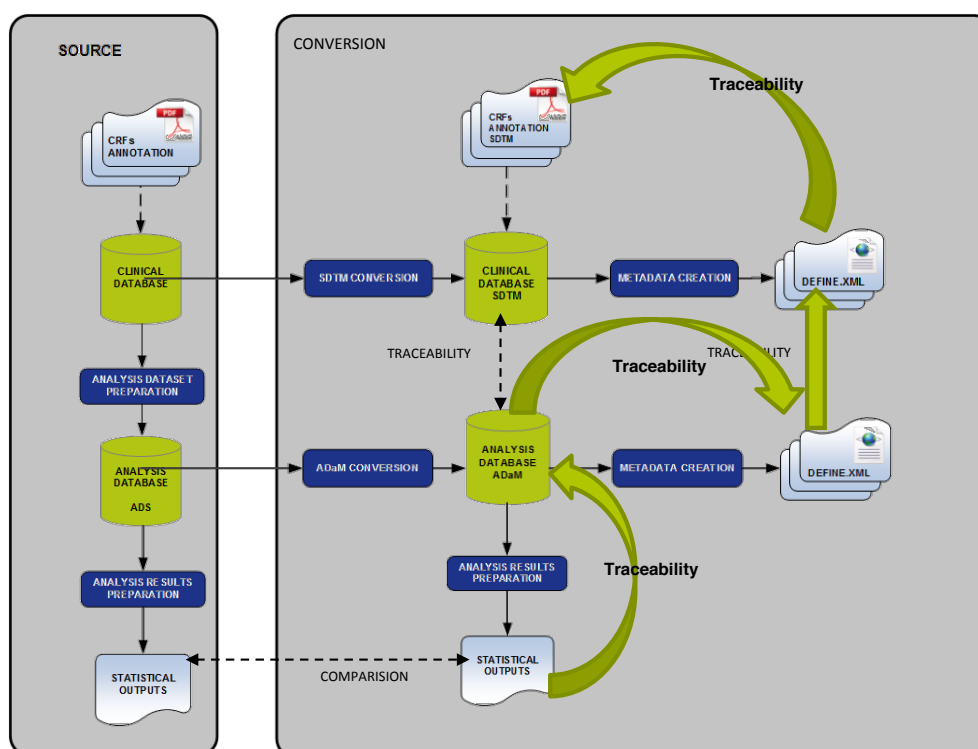
ADaM Conversions Two approaches: Linear and Parallel



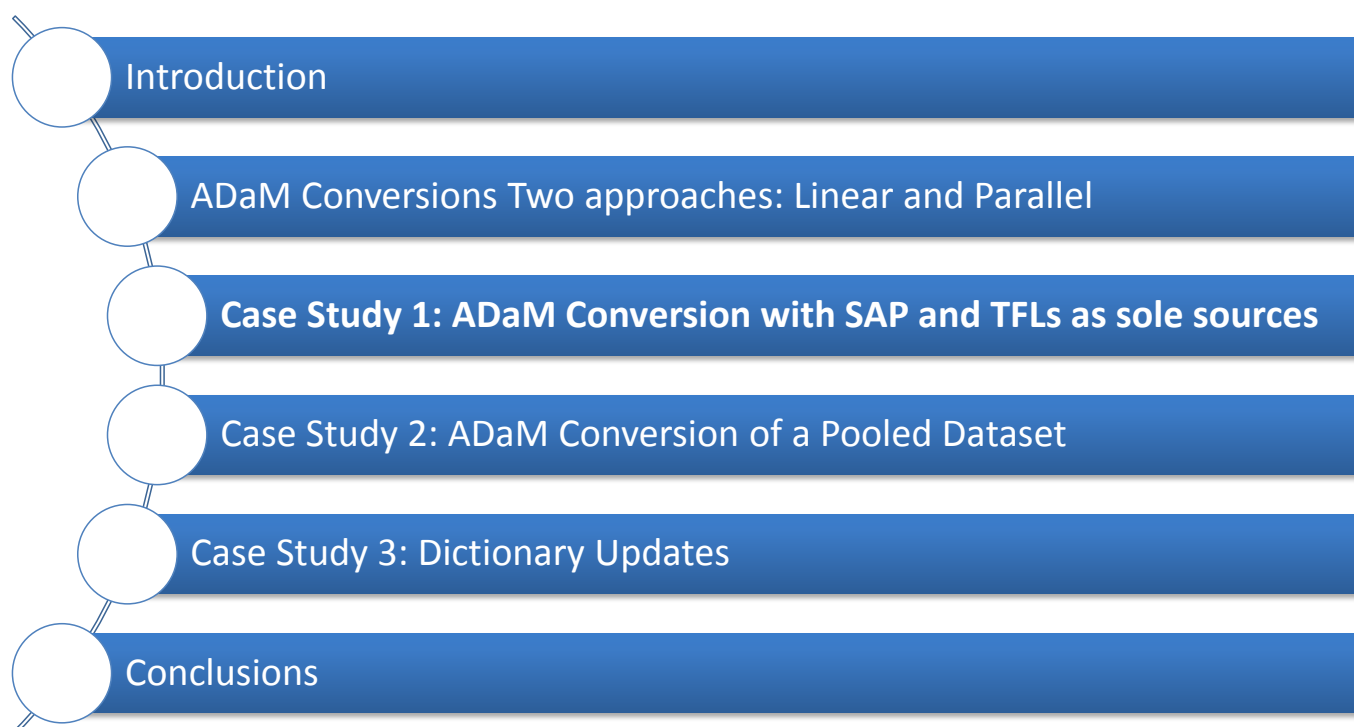
ADaM Conversions Two approaches: Linear and **Parallel**



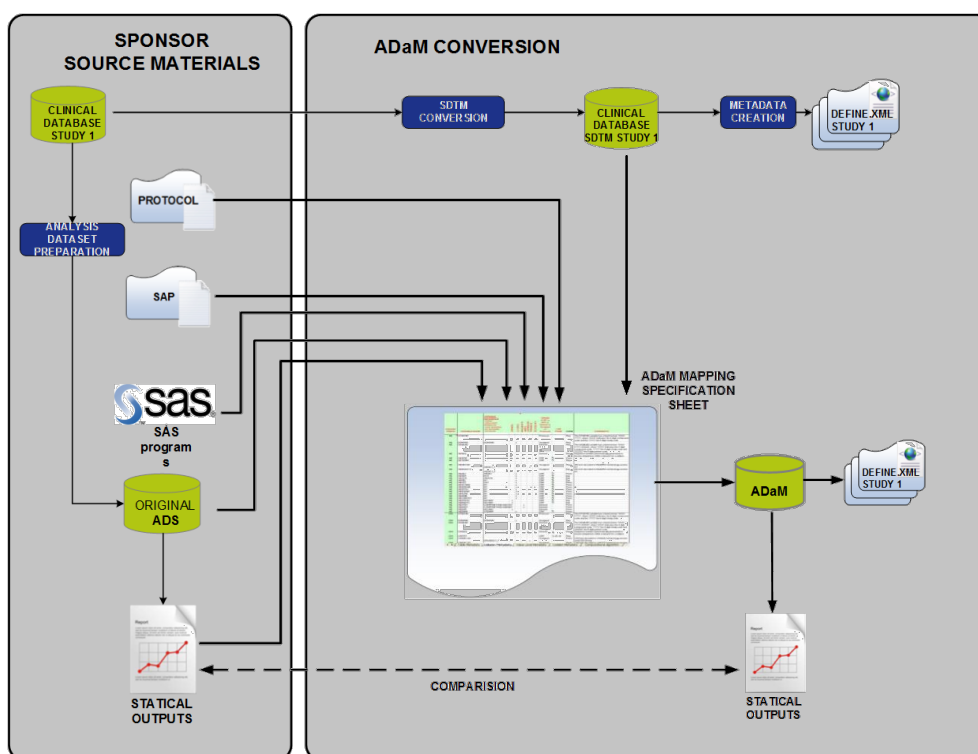
ADaM Conversions Two approaches: Linear and **Parallel**



ADaM Conversion: The Good, the Bad and the Ugly



Case Study 1: ADaM Conversion with SAP and TFLs as sole Sources



Case Study 1: ADaM Conversion with SAP and TFLs as sole Sources



- **Good Traceability** between the statistical outputs, the ADaM datasets and the SDTM domains
- Since BDLS was responsible for the **SDTM Conversion**, issues with SDTM domains were resolved rapidly
- Having the ADS in ADaM structure facilitated the addition of this legacy study as part of an **Analysis Pooled Database**

Case Study 1: ADaM Conversion with SAP and TFLs as sole Sources



**T
H
E
B
A
D**



- Hidden differences with SDTM found during validation:
 - Different Coding Dictionary Versions in SDTM database
 - For consistency with newer studies, some categorical variables were updated with new combination of values
- Due to SDTM standards restrictions some values used for analysis were not available in the SDTM datasets
 - Key variables in SDTM domains cannot be empty
 - In the initial analysis missing values were summarized in a table, these values were not available anymore in the SDTM datasets

Case Study 1: ADaM Conversion with SAP and TFLs as sole Sources

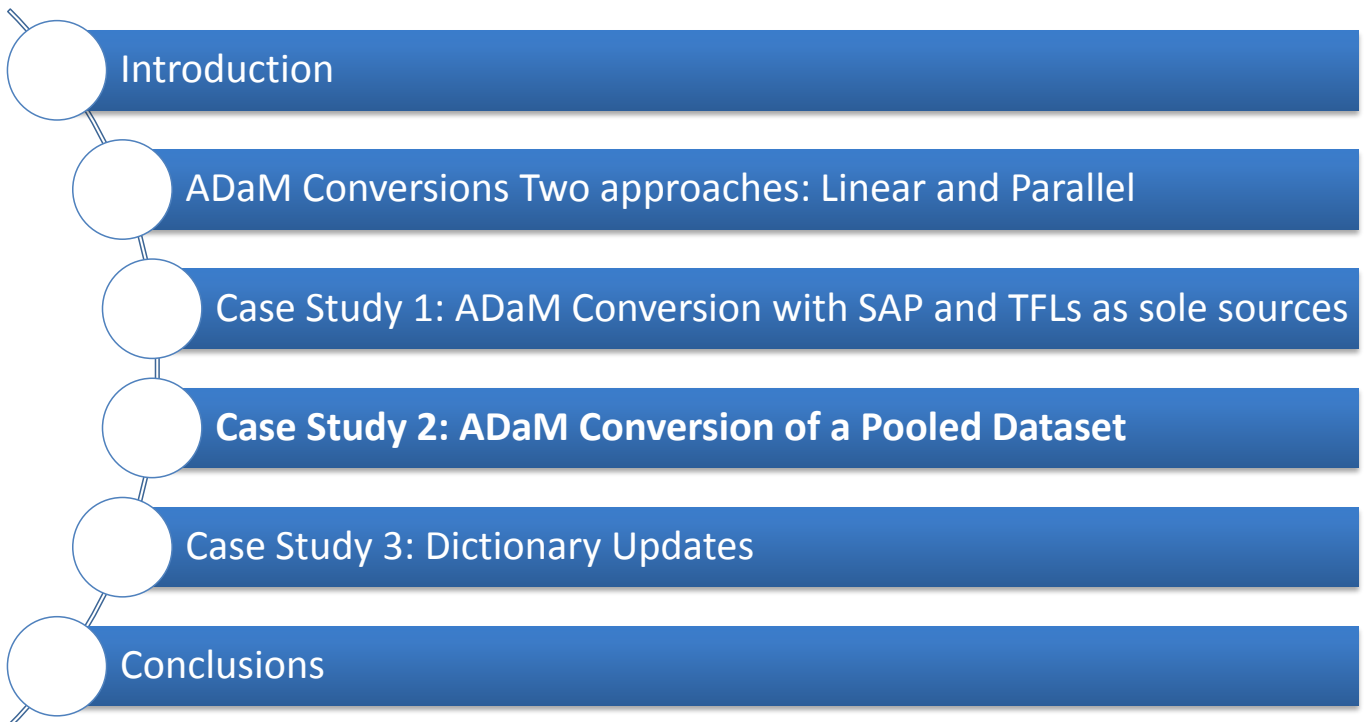


T
H
E
U
G
L
Y

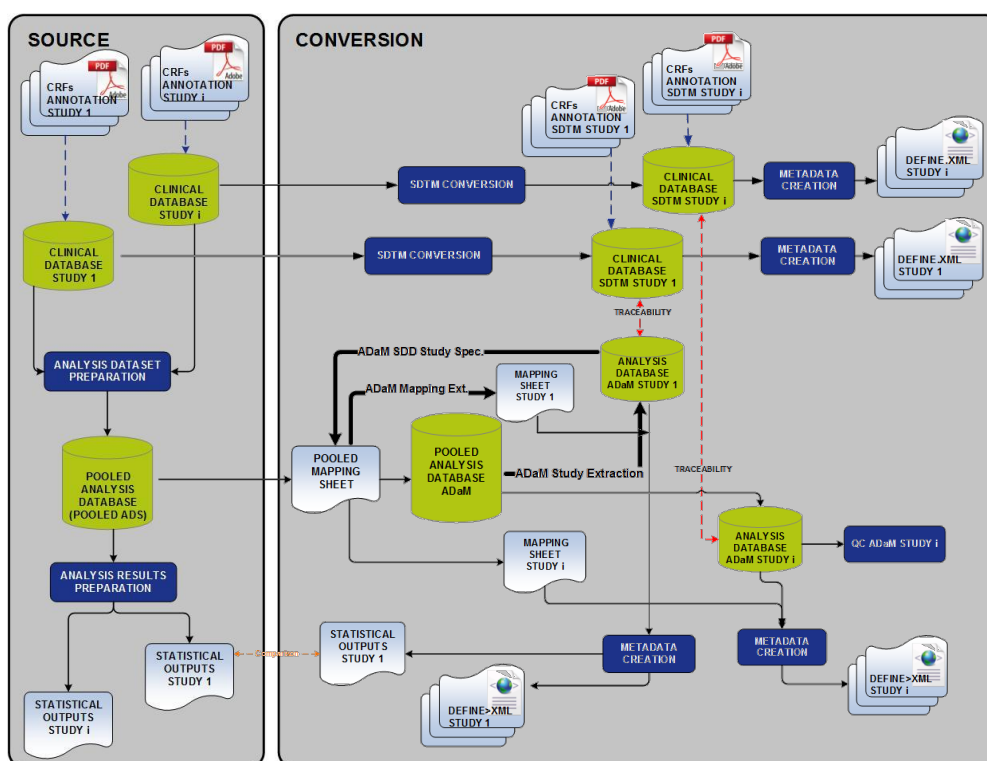


- **ADaM specifications** created from SAP and protocol
 - Rules specified in Protocol and SAP were differently interpreted in the same way as in the original analysis
 - Multiple solutions were tried before having the proper result
- **QC: manual comparisons were needed**
 - **ADS** not always available
 - **Listings** in pdf format
 - Resolving all these issues was time consuming

ADaM Conversion: The Good, the Bad and the Ugly

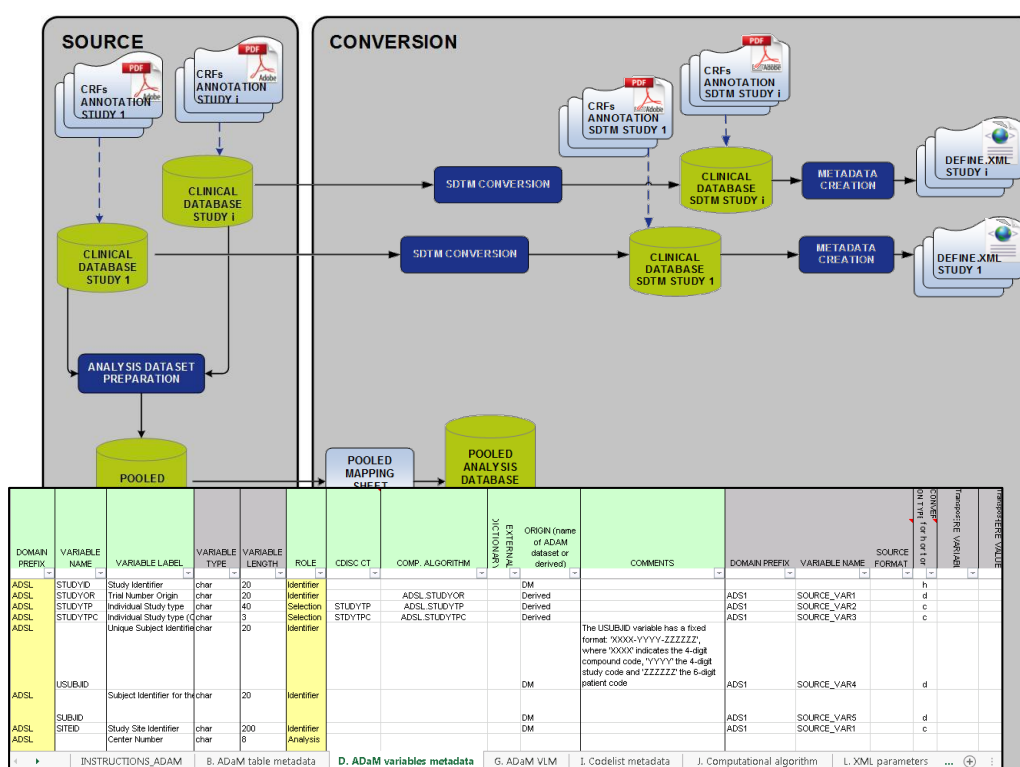


Case Study 2: ADaM Conversion of a Pooled Dataset



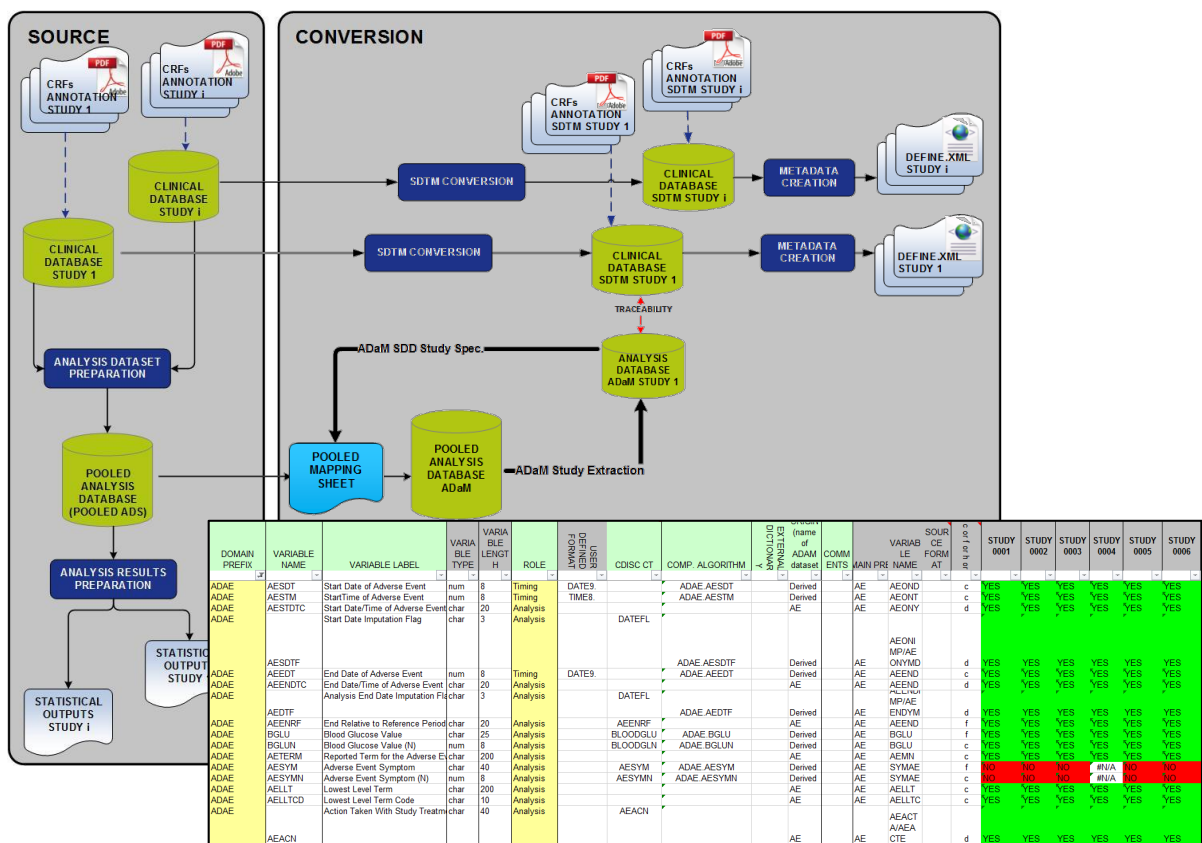
Case Study 2:

ADaM Conversion of a Pooled Dataset

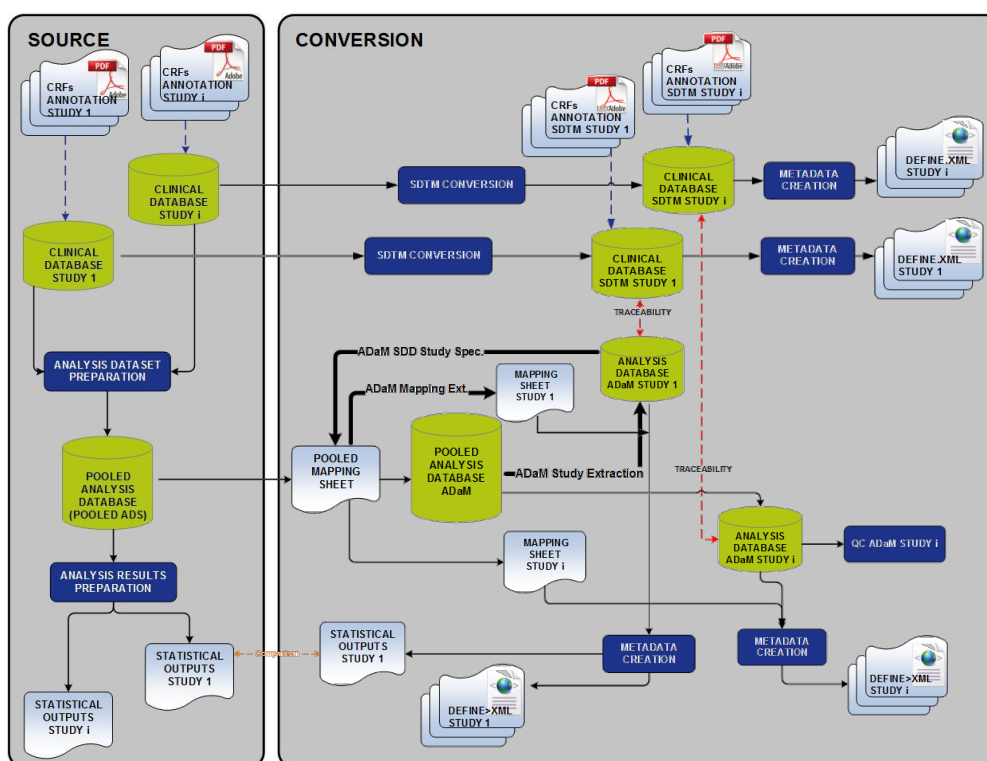


Case Study 2:

ADaM Conversion of a Pooled Dataset



Case Study 2: ADaM Conversion of a Pooled Dataset



Case Study 2:

ADaM Conversion of a Pooled Dataset



T
H
E

G
O
O
D



- BDLS was in charge of the SDTM and ADaM conversion, this **facilitated the communication** between the two teams and accelerated the problem solving
- With a **mature Pooled ADaM conversion process**, data extraction for each study is fast, and consistency checks across studies are no longer needed
- **Codelist were re-usable** in ADaM due to its consistency across studies
- A **comment file** with all questions from and to the sponsor, facilitates the tracking of all answers and decisions

Case Study 2:

ADaM Conversion of a Pooled Dataset



T
H
E

B
A
D



- Not well documented **Derivation Rules** caused difficulties to define the traceability
- **Studies were added into the pooled database** at different time points, while derivations were not properly updated
- **SDTM limitations** affected the ADaM conversion
 - some raw variables were not available in SDTM, but were mentioned in ADS specifications
 - Specifications had to be redefined in cooperation with the sponsor
- Due to traceability between SDTM/ADaM, **timelines** for ADaM datasets were influenced by SDTM timelines

Case Study 2:

ADaM Conversion of a Pooled Dataset



T
H
E

U
G
L
Y



- **Structure of the pooled original ADS datasets** changed with the inclusion of new studies
 - Impact on the pooled mapping, pooled datasets and extractions
- **Source variables** which had some manipulation during the conversion (e.g. variable derived, with a format or transposed) were **sensitive to changes in its content** at each transfer
 - New compare tools to check changes in dataset structure between the original ADS received

Case Study 2:

ADaM Conversion of a Pooled Dataset



- BDLS developed **new compare tools** to check changes in dataset structure between the original ADS received, like number of records, empty, new, and missing variables that were present before
- Specifications** were **compared**: results of a compare were included in the communication file to the sponsor

	B	C	D	E	F	H	I	J
	VARIABLE	DESCRIPTION	TYE	compare_flg	new_domain_flag	B&DLS comments	Sponsor comments	Updates
1	SOURCE_VAR_1	AESI flag-Angioedem num	ADDED	NO				
2	SOURCE_VAR_2	AESI flag-Embolic an num	ADDED	NO				
3	SOURCE_VAR_3	AESI flag-Hypoglycaenum	ADDED	NO				
4	SOURCE_VAR_4	AESI flag-Hypersensi num	MODIFIED	YES				
5	SOURCE_VAR_5	AESI flag-Increased u num	ADDED	YES				
6	SOURCE_VAR_6	Investigator Special Ir num	REMOVED	NO				
7	SOURCE_VAR_7	Investigator Special Ir num	REMOVED	NO				
8	SOURCE_VAR_8	Investigator Special Ir num	REMOVED	NO				
9	SOURCE_VAR_9	Investigator Special Ir num	ADDED	NO				
10	SOURCE_VAR_10	Investigator Special Ir num	ADDED	NO				
11	SOURCE_VAR_11	Investigator Special Ir num	ADDED	NO				
12	SOURCE_VAR_12	Investigator Special Ir num	ADDED	NO				
13	SOURCE_VAR_13	Investigator Special Ir num	ADDED	NO				
14	SOURCE_VAR_13	Investigator Special Ir num	ADDED	NO				
Issues & updates post-release Source data issues specs_comp Compare_tool SDTM Source not mapped Review cycle								

Case Study 2:

ADaM Conversion of a Pooled Dataset



- ADS content was compared:**

Results from content compare for variables being “derived”, “with a CL applied”, “transposed” during ADaM conversion, were included in the communication file to the sponsor

The below compare results are from comparing the content of variables in old transfer against the new transfer. This compare is done ONLY for variables that are 'derived' or have a format applied while being mapped into ADaM. Any change in the source implies a change in ADaM derivation rule, e.g. transpose rule, or code list.

A	B	C	D	E	F	G	H
studyid	DOMAIN	variable	comment	content	B&DLS Comment	Sponsor	Updates Comments
0001	ADS11	SRC_VAR_5	new variable to transpose present in new database	Creatinine >= 1.5* ref. sample and > ULN	we will include a new criteria for this value. Please confirm the value is correct	correct	Pooled mapping updated
0001	ADS2	SRC_VAR_7	Domain not present in new database	Creatinine >= 2*baseline and > ULN	we will include a new criteria for this value. Please confirm the value is correct	correct	not included
0001	ADS9	SRC_VAR_5	Value not present in code list metadata but present in source data	TS_CV	this is a new set of population flag, please confirm if we need to add it	Please ignore	not included

Issues & updates post-release | Source data issues | specs_comp | **Compare tool** | SDTM Source not mapped | Review cycles | Consistency checks | Def ... (+) |

Case Study 2:

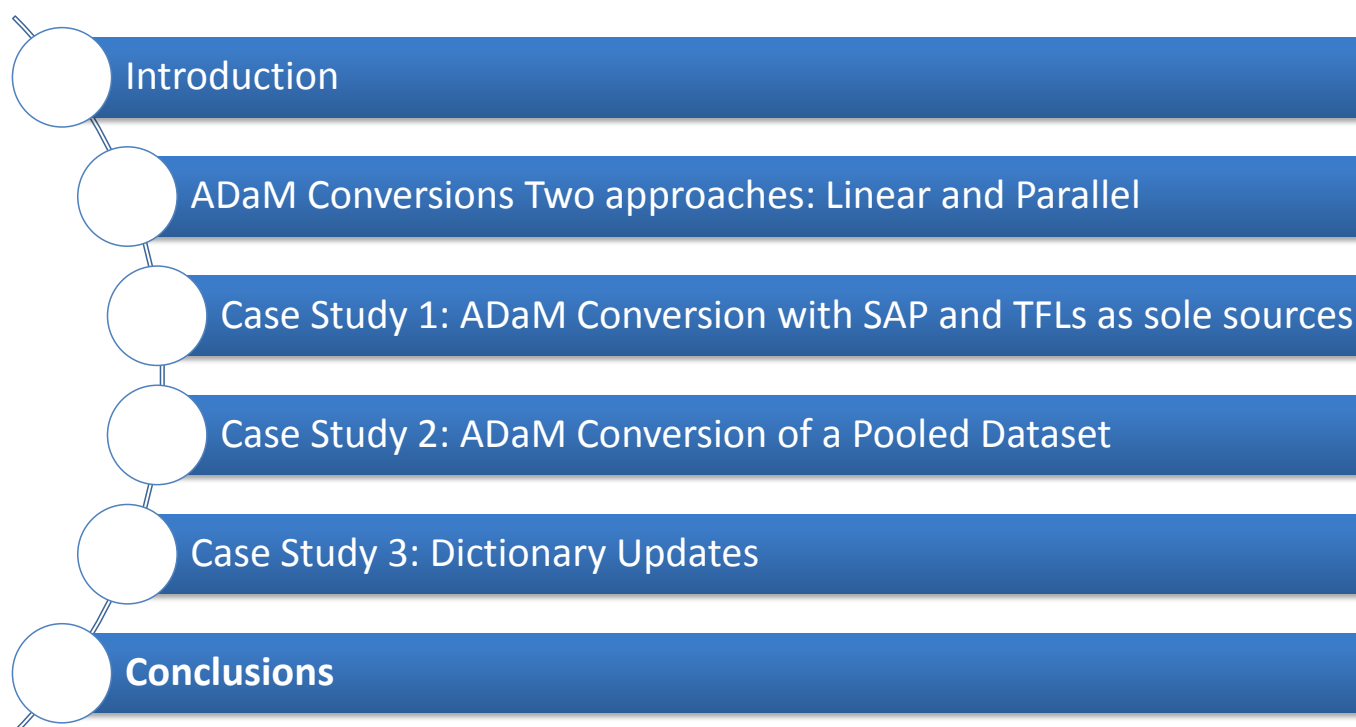
ADaM Conversion of a Pooled Dataset



- To ensure traceability with SDTM, a good knowledge of **what was mapped from RAW into SDTM** and in which study, was essential
- List of RAW variables per study** per domain from each SDTM metadata, specifying if the variable was mapped into SDTM, and if it was empty

STUDY	VIEWNAME	VARNAME	LABEL	VARTYPE	FORMATN	MAPPED	CONFIRMED	EMPTY
RAW_Study1	AE	AEI	AE indicator	num	YN1F.		x	X
RAW_Study1	AE	AEONTM	AE onset time	num	TIME5.	X	x	
RAW_Study1	AE	AEENDC	AE end date continued	char	\$CONT1F.	X	x	
RAW_Study1	AE	AEOUT	AE outcome	num	AEOUT1F.	X	x	
RAW_Study1	AE	AEREL	AE drug relationship	num	YN1F.	X	x	
RAW_Study1	AE	AEONDT	AE onset date	num	DATE9.	X	x	
RAW_Study1	AE	AEENDDT	AE end date	num	DATE9.	X	x	
RAW_Study1	AE	AELLT	AE lowest level term	char	\$200.		x	
RAW_Study1	AE	AELLTCD	AE lowest level term code	char	\$10.		x	
RAW_Study2	AE	AEI	AE indicator	num	YN1F.		x	X
RAW_Study2	AE	AEONTM	AE onset time	num	TIME5.	X	x	
RAW_Study2	AE	AEENDC	AE end date continued	char	\$CONT1F.	X	x	
RAW_Study2	AE	AEOUT	AE outcome	num	AEOUT1F.	X	x	
RAW_Study2	AE	AEREL	AE drug relationship	num	YN1F.	X	x	
RAW_Study2	AE	AEONDT	AE onset date	num	DATE9.	X	x	
RAW_Study2	AE	AEENDDT	AE end date	num	DATE9.	X	x	
RAW_Study2	AE	AELLT	AE lowest level term	char	\$200.		x	
RAW_Study2	AE	AELLTCD	AE lowest level term code	char	\$10.		x	

ADaM Conversion: The good, the bad and the ugly



Conclusions



- **Parallel approach:**
 - ensured reproduction of TFLs
 - consistency between studies within a pooled database
- **Linear approach:**
 - SDTM/ADaM traceability is demonstrated
 - Bigger risk of not reproducing identical results to the initial analysis
- **Source pooled ADS** should be stable and **specifications** should be clear
- SDTM datasets & pooled ADaM datasets should be **finalized** before the individual study extractions can start

Conclusions



- Dictionary update: SDTM and ADaM uses the **same dictionary version**. Specific ADaM variables are available
- **Good communication** is needed between SDTM, ADaM team members and the sponsor, during the whole process
- **Sponsor's feedback** on draft mapping and draft ADaM datasets is crucial to address all details upfront
- All decisions can be easily tracked if they are properly documented in a **communication file**



Thank you for your attention.

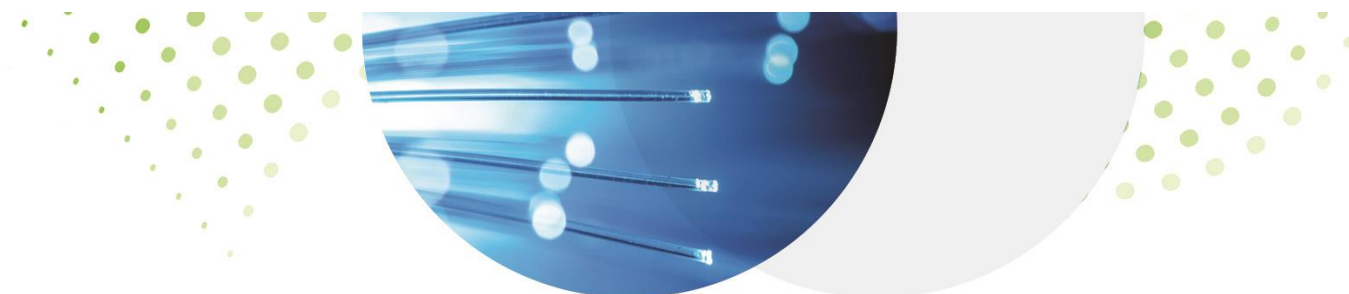
Join us for a coffee at our stand.

London, United Kingdom, 13 OCT 2014



Restricted © Business & Decision Life Sciences 2014 All rights reserved.





Carla Santillán | Project Manager | Statistical Projects carla.santillan@businessdecision.com
Jessica Minkue Mi Edou | Statistical Programmer | jessica.minkue@businessdecision.com

Business & Decision Life Sciences
Sint-Lambertusstraat 141 rue Saint-Lambert
B-1200 Brussels
T: +32 2 774 11 00
F: +32 2 774 11 99
lifesciences@businessdecision.com
<http://www.businessdecision-lifesciences.com/>