



SD16

ADaM Intermediate Dataset

How to improve your analysis traceability

Angelo Tinazzi, Teresa Curto, Ashish Aggarwal
Cytel Inc.



- Introduction
- Intermediate ADaM Datasets
- Intermediate ADaM Datasets – An Example
- Conclusions

Introduction

Introducing Traceability

Introduction

What is Traceability?

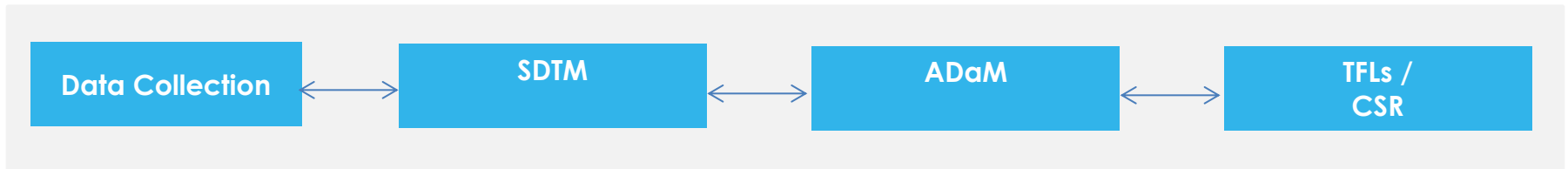
Source	Definition
Wikipedia	Traceability is the ability to verify the history, location, or application of an item by means of documented recorded identification Ref. "Glossary," ASME Boiler and Pressure Vessel Code, Section III, Article NCA-9000
GCP	All clinical trial information should be recorded, handled, and stored in a way that allows its accurate reporting, interpretation and verification Ref. ICH E6 GCP Section 2.10 "GCP Principles"
CDISC	<i>The property that enables the understanding of the data's lineage and/or the relationship between an element and its predecessor(s).</i> Ref. CDISC ADaM Ig 1.1
FDA	<i>An important component of a regulatory review is an understanding of the provenance of the data (i.e., traceability of the sponsor's results back to the CRF data).</i> <i>Traceability permits an understanding of the relationships between the analysis results, analysis datasets, tabulation datasets, and source data.</i> Ref. FDA Study Data Technical Conformance Guide (SDTCG)

Introduction

What is Traceability?

Lost in Traceability

Provide a clear , traceable pathway from the primary collection documents (e.g. CRF, Diary, External Data) to the raw datasets (SDTM) to the analysis datasets (ADaM)* to your TFL (Table Figures and Listings) to your CSR (Clinical Study Report)



* Optimizing Your Study Data Submissions to FDA – Updates from CDER and CBER, July 2017, <http://sbiaevents.com/files/study-data-071217.pdf>

Introduction

Data-Point Traceability in ADaM

Points directly to the specific predecessor record(s) and should be implemented if practical and feasible.

ADAE

USUBJID	AESEQ	AETERM	AESTDTC
██████-005-1001-13	4	HYPERTENSION	2017-08-22
██████-005-1001-13	5	BIOCHEMICAL PREGNANCY	2017-08-23
██████-005-1003-04	1	DISCOMFORT AFTER OOCYTE PICK UP	2017-09-22



Where are my data coming from?

AE

STUDYID	DOMAIN	USUBJID	AESEQ	AESPID	AETERM
██████-005	AE	██████-005-1001-13	4	004	HYPERTENSION
██████-005	AE	██████-005-1001-13	5	005	BIOCHEMICAL PREGNANCY
██████-005	AE	██████-005-1003-04	1	001	DISCOMFORT AFTER OOCYTE PICK UP



Introduction

Metadata Traceability in ADaM

- **Facilitates the understanding** of the relationship of the analysis variable to its source dataset(s) and variable(s).
- Metadata traceability is also used to establish the **relationship between an analysis result and ADaM dataset(s)**.
- Make sure you have a **good define.xml** e.g. derivations clearly described

Parameter Value List - ADTTE [AVAL]

Variable	Where	Type	Length / Display Format	Controlled Terms or Format	Source/Derivation/Comment
AVAL	PARAMCD = "TTIBG360" (Time to Increased Blood Glucose Levels (Hours) 1-360)	integer	8		Derived: Time to the earliest significantly increased blood glucose level in hours was derived as the time from administration of study treatment (ADSL.TR01SDT) to the first blood glucose level (AVAL when ADBGL.PARAMCD = 'BGLAVGHR') that is ≥ 2 SDs above the 72-hour baseline average blood glucose level (AVAL when ADBGL.PARAMCD = 'BGLAVGDL' and AVISITN = 2 and BASETYPE = '72HRS'). AVAL is the relative hour (ATPTN) of first increased blood glucose level till 360 Hours. Patients without a ≥ 2 SD increase during the time period was censored at the end of the period.

Analysis Data Reviewer's Guide

Sponsor
Study XXXX-2015-010
Version 1

Study XXXX-2015-010

Analysis Data Reviewer's Guide

Analysis Data Reviewer's Guide

Contents

1. Introduction.....	4
1.1 Purpose.....	4
1.2 Acronyms.....	4
1.3 Study Data Standards and Dictionary Inventory.....	4
1.4 Source Data Used for Analysis Dataset Creation.....	5
2. Protocol Description.....	6
2.1 Protocol Number and Title.....	6
2.2 Protocol Design in Relation to ADaM Concepts.....	6
3. Analysis Considerations Related to Multiple Analysis Datasets.....	7
3.1 Comparison of SDTM and ADaM Content.....	7
3.2 Core Variables.....	7
3.3 Treatment Variables.....	8
3.4 Subject Issues that Require Special Analysis Rules.....	8
3.5 Use of Visit Windowing, Unscheduled Visits, and Record Selection.....	8
3.6 Imputation/Derivation Methods.....	9
4. Analysis Data Creation and Processing Issues.....	9
4.1 Split Datasets.....	9
4.2 Data Dependencies.....	9
4.3 Intermediate Datasets.....	10
4.4 Variable Conventions.....	10
5. Analysis Dataset Descriptions.....	11
5.1 Overview.....	11
5.2 Analysis Datasets.....	11
5.2.1 ADSL – Subject Level Analysis Dataset.....	12
5.2.1 ADOA – Osteoarthritis History AD.....	12
5.2.3 ADGLRLHR – Intermediate Blood Glucose AD.....	13
5.2.4 ADBGL – Blood Glucose AD.....	13
5.2.5 ADBGLAUE – Blood Glucose Area Under the Effect (AUE) AD.....	14
5.2.6 ADBGCAT – Blood Glucose Categories AD.....	15
5.2.7 ADTTE – Time to Event AD.....	16
5.2.8 ADAE – Adverse Events AD.....	17
5.2.9 ADAESUM – Adverse Events Summary AD.....	17

PDF created 2017-01-19

Page 2 of 30

Introduction

Traceability=Transparency



When traceability is successfully implemented, it is possible to identify:

- Information that **exists in the submitted SDTM data**
- Information that is **derived or imputed** within the ADaM dataset
- The method used to create **derived or imputed** data
- **Information used for analyses**, in contrast to **information that is not used for analyses** yet is included to support traceability or future analysis

Ref. ADaM Ig 1.1 Section 2.2

Intermediate ADaM Dataset

Improve your Traceability

Intermediate ADaM Dataset From ADaM Ig 1.1



- **Very complex derivations** may require the creation of intermediate analysis datasets.
- In these situations, traceability may be accomplished by **submitting those intermediate analysis datasets** along with their associated metadata.
- Traceability would then **involve several steps**.
- The analysis results would be linked by **appropriate metadata** to the data which supports the analytical procedure, **those data would be linked to the intermediate analysis data**, and the **intermediate data would in turn be linked to the source SDTM data**

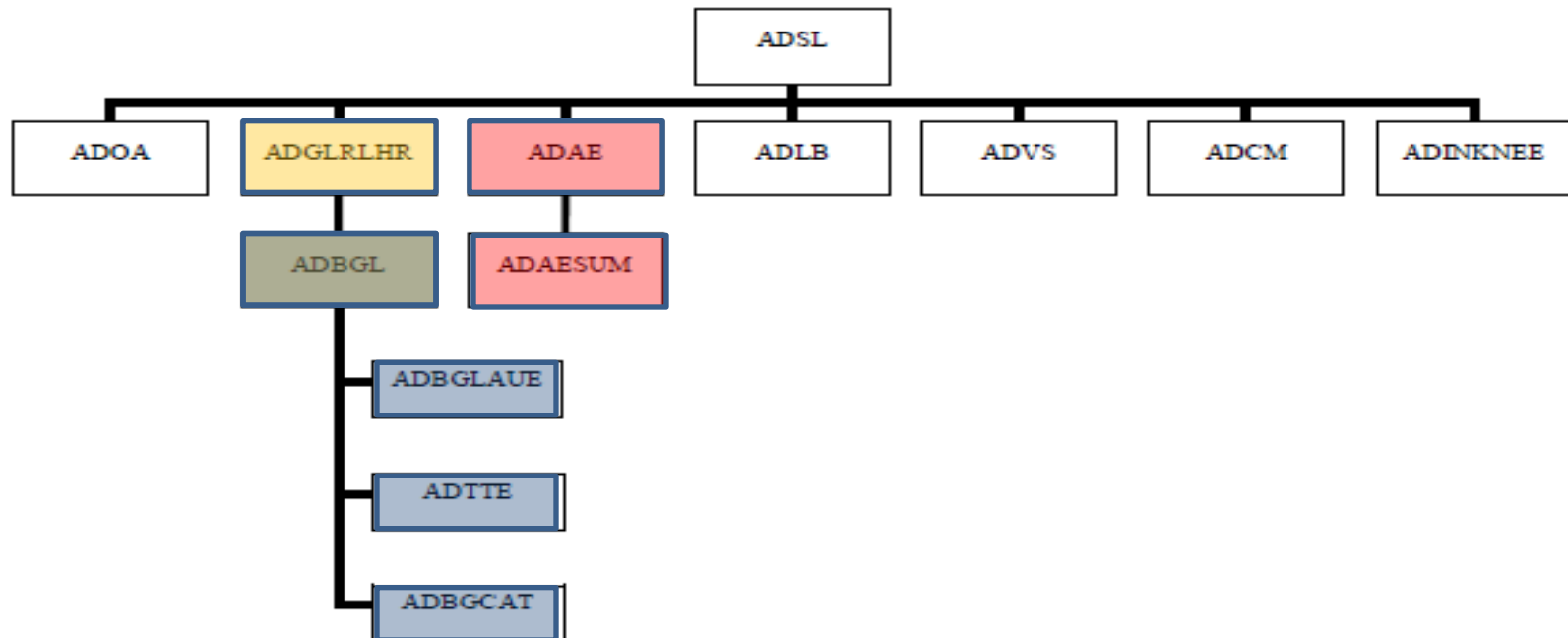
4.2 Data Dependencies

ADSL was used in the creation of all other analysis datasets.

ADGLRLHR was used in the creation of ADBGL.

ADBGL was used in the creation of ADBGCAT, ADBGLAUE and ADTTE.

ADAE was used in the creation of ADAESUM.



Intermediate ADaM Dataset Example for TTE



End-point Progression Free Survival where the event is either tumor progression or death whichever come first, with event censored when other therapies are taken

ADTTEDAT

Row	USUBJID	ASEQ	ASTDT	ASTDY	PARAMCD	PARAM	AVALC	ANL01 FL	SRCDOM	SRCVAR	SRCSEQ
1	S1-01-01	1	01DEC2011	1	DISPOSI	Disposition	RANDOMIZED	Y	DS	DSSTDTC	1
2	S1-01-01	2	22DEC2011	22	ASSESS	Tumor Assessment	SD	Y	RS	RSDTC	1
3	S1-01-01	3	10JAN2012	41	ASSESS	Tumor Assessment	SD	Y	RS	RSDTC	2
4	S1-01-01	4	31JAN2012	62	ASSESS	Tumor Assessment	PD	Y	RS	RSDTC	3
5	S1-01-01	5	02FEB2012	64	DEATH	Death	YES	Y	DM	DTHDTC	
6	S1-01-02	1	01JAN2012	1	DISPOSIT	Disposition	RANDOMIZED	Y	DS	DSSTDTC	1
7	S1-01-02	2	20JAN2012	20	ASSESS	Tumor Assessment	SD	Y	RS	RSDTC	1
8	S1-01-02	3	10FEB2012	41	ADDEVENT	Additional Event	PROHIBITED MEDICATION		CM	CMSTDTC	14
9	S1-01-02	4	28FEB2012	59	ASSESS	Tumor Assessment	PD		RS	RSDTC	2

EVENT

CENSORED

ADTTEPFS

Row	USUBJID	PARAMCD	PARAM	ASTDT	ADT	AVAL	CNSR	SRCDOM	SRCVAR	SRCSEQ
1	S1-01-01	PFS	Progression Free Survival (Days)	01DEC2011	31JAN2012	62	0	ADTTEDAT	AVAL	4
2	S1-01-02	PFS	Progression Free Survival (Days)	01JAN2012	20JAN2012	20	1	ADTTEDAT	AVAL	2

Intermediate ADaM: an example

From individual measurements to
different layers of aggregation

Intermediate ADaM: an example

The “Scenario” – Mini Protocol Synopsis



- Study: Phase II
- Design: Double-Blind, Randomized
- Endpoint: effects of Treatment A vs. Treatment B on blood glucose levels.
- Blood glucose collected continually using continuous glucose monitor (CGM)
- Measurements taken in 5 minute increments.

Intermediate ADaM: a use case

Definition: Average daily blood glucose



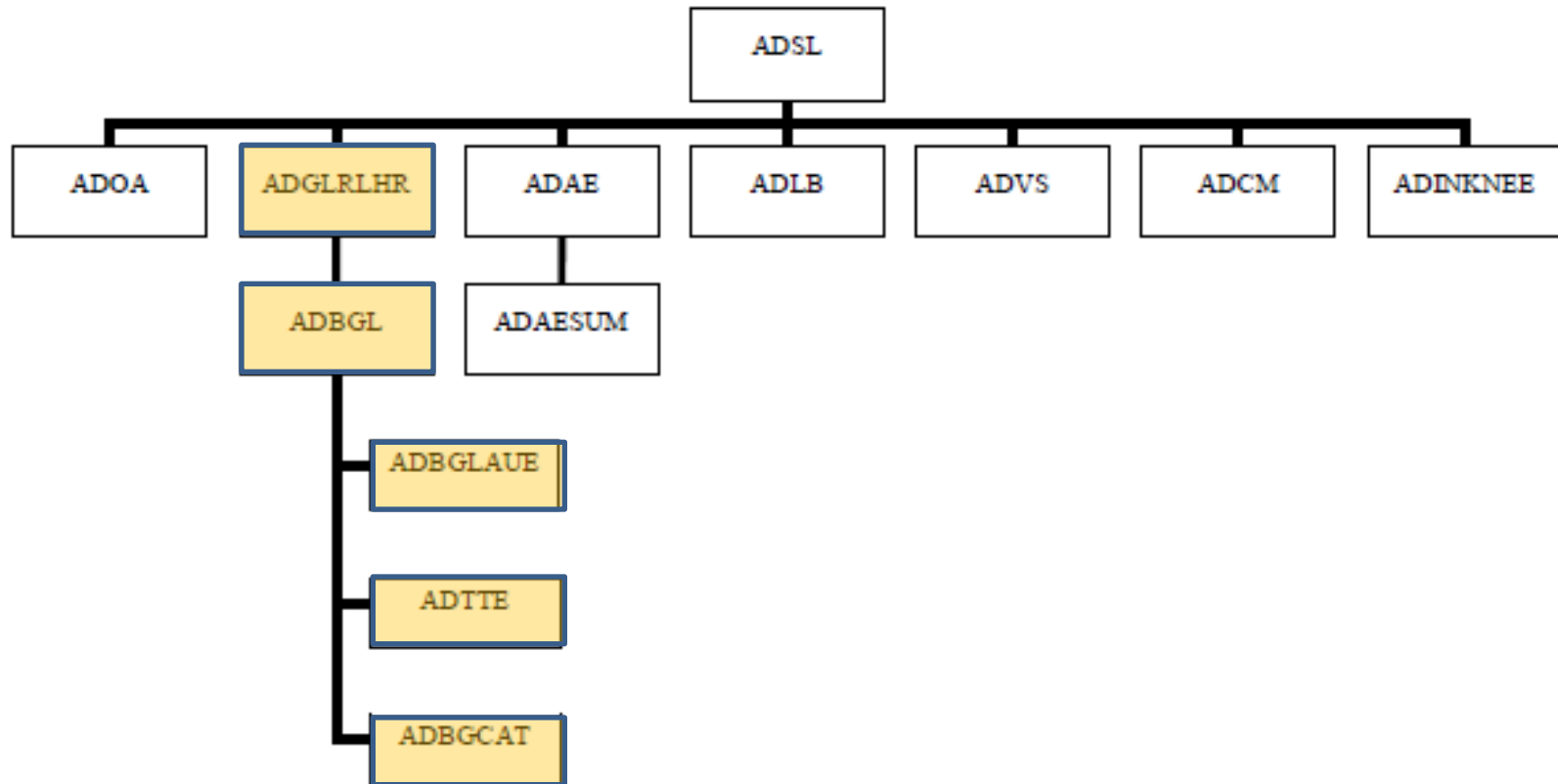
From the SAP:

▲ 4.8.1. Average daily blood glucose

Average daily blood glucose will be calculated in one-hour increments from time of dosing to Day 15 (Hour 360). Average blood glucose is first calculated for each hour by subject using all non-missing values. Then the mean of each of these hourly averages is taken to calculate the mean for each subject within each specified interval. The denominator is the number of non-missing records for each interval.

Intermediate ADaM: a use case

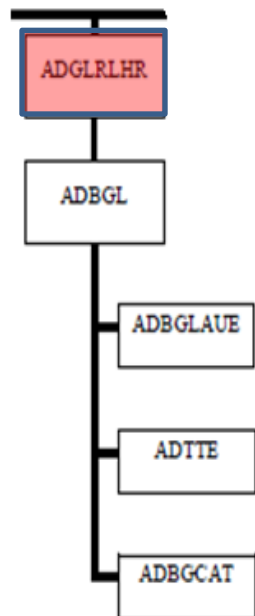
Definition: Average daily blood glucose



Intermediate ADaM: a use case

Step 1 : Windowing time-point and visit

ADGLRLHR - Intermediate Blood Glucose AD



- Blood Glucose automatically measured by a device
- No visits / time-points reference
- A windowing method was defined in the SAP

Dose Time	Relative Hour (ATPT)	Start Time	End Time
	-4	10:00:01	11:00:00
	-3	11:00:01	12:00:00
	-2	12:00:01	13:00:00
	-1	13:00:01	13:14:59
13:15	1	13:15:00	14:00:00
	2	14:00:01	15:00:00
	3	15:00:01	16:00:00
	4	16:00:01	17:00:00

Relative Hour Derivation Example

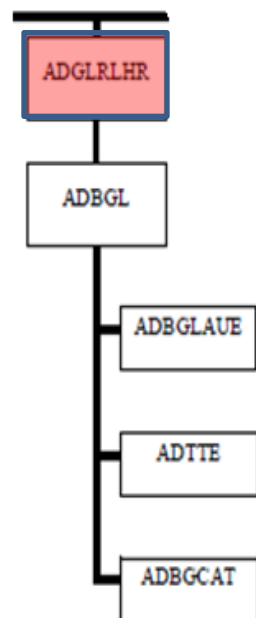
Visit (AVISIT)	Start	End
Baseline	Hour -72	Hour -1
Day 1	Hour 1	Hour 24
Day 2	Hour 25	Hour 48
Day 3	Hour 49	Hour 72
Day 4	Hour 73	Hour 96
Day 5	Hour 97	Hour 120
Day 6	Hour 121	Hour 144
Day 7	Hour 145	Hour 168
Day 8	Hour 169	Hour 192
Day 9	Hour 193	Hour 216
Day 10	Hour 217	Hour 240
Day 11	Hour 241	Hour 264
Day 12	Hour 265	Hour 288
Day 13	Hour 289	Hour 312
Day 14	Hour 313	Hour 336
Day 15	Hour 337	Hour 360
Day 1-3	Hour 1	Hour 72

Visit Windows

Intermediate ADaM: a use case

Step 1 : Windowing time-point and visit

ADGLRLHR - Intermediate Blood Glucose AD



PARAM	PARAMCD	AVAL	ADT	ATM	AVISIT	ATPT	ATPTN	ANL01FL	ZGSEQ
Blood Glucose Level(mg/dL)	BGL	171	02MAY2016	14:16		Hour -73	-73		1171
Blood Glucose Level(mg/dL)	BGL	182	02MAY2016	14:21		Hour -73	-73		1172
Blood Glucose Level(mg/dL)	BGL	195	02MAY2016	14:26		Hour -73	-73		1173
Blood Glucose Level(mg/dL)	BGL	201	02MAY2016	14:31		Hour -73	-73		1174
Blood Glucose Level(mg/dL)	BGL	200	02MAY2016	14:36		Hour -73	-73		1175
Blood Glucose Level(mg/dL)	BGL	199	02MAY2016	14:41		Hour -73	-73		1176
Blood Glucose Level(mg/dL)	BGL	198	02MAY2016	14:46		Hour -73	-73		1177
Blood Glucose Level(mg/dL)	BGL	190	02MAY2016	14:51		Hour -73	-73		1178
Blood Glucose Level(mg/dL)	BGL	183	02MAY2016	14:56		Hour -73	-73		1179
Blood Glucose Level(mg/dL)	BGL	174	02MAY2016	15:01	Baseline	Hour -72	-72	Y	1180
Blood Glucose Level(mg/dL)	BGL	165	02MAY2016	15:06	Baseline	Hour -72	-72	Y	1181
Blood Glucose Level(mg/dL)	BGL	162	02MAY2016	15:11	Baseline	Hour -72	-72	Y	1182
Blood Glucose Level(mg/dL)	BGL	158	02MAY2016	15:16	Baseline	Hour -72	-72	Y	1183

ATPT and AVISIT as per
planned analysis
windowing

Visit (AVISIT)	Start	End
Baseline	Hour -72	Hour -1

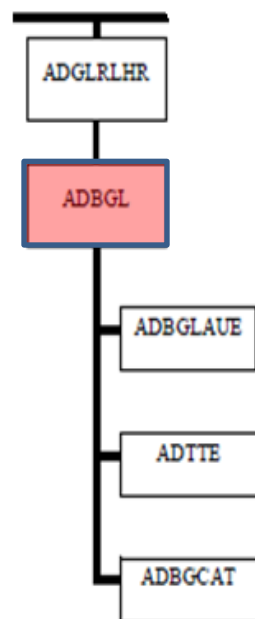
Data-Point
Traceability
to SDTM

Assessments prior to
Baseline (prior to hour
-72) will be not analyzed

Intermediate ADaM: a use case

Step 2: Hourly and daily averages

ADBGL – Blood Glucose AD



- Change from Baseline analysis by time-point for hourly average (PARAMCD=BGLAVGHR)
- Change from Baseline analysis by time-point for daily average (PARAMCD=BGLAVGDL)

AVISIT	ATPT	ATPTN	PARAM	PARAMCD	AVAL	AVALCAT1	BASE	CHG	ASEQ
Day 14	Hour 330	330	Hourly Avg. Blood Glucose Level (mg/dL)	BGLAVGHR	142.1		.		387
Day 14	Hour 331	331	Hourly Avg. Blood Glucose Level (mg/dL)	BGLAVGHR	136.9		.		388
Day 14	Hour 332	332	Hourly Avg. Blood Glucose Level (mg/dL)	BGLAVGHR	143.4		.		389
Day 14	Hour 333	333	Hourly Avg. Blood Glucose Level (mg/dL)	BGLAVGHR	174.3		.		390
Baseline		.	Daily Avg. Blood Glucose Level (mg/dL)	BGLAVGDL	195.8	Blood glucose >= 180.1	195.8		393
Day 1		.	Daily Avg. Blood Glucose Level (mg/dL)	BGLAVGDL	212.9	Blood glucose >= 180.1	195.8	17.1	396
Day 2		.	Daily Avg. Blood Glucose Level (mg/dL)	BGLAVGDL	207.6	Blood glucose >= 180.1	195.8	11.9	399
Day 3		.	Daily Avg. Blood Glucose Level (mg/dL)	BGLAVGDL	203.5	Blood glucose >= 180.1	195.8	7.8	402

Metadata Traceability through define.xml

Created for later data-point traceability when ADBGL will serve as 'intermediate' ADaM

AVAL	<u>PARAMCD</u> = "BGLAVGDL" (Daily Avg. Blood Glucose Level (mg/dL))	float	8	Derived: Daily Average Blood Glucose Level (mg/dL) defined as average of hourly average blood glucose (ADBGL.BGLAVGHR).
------	--	-------	---	--

Intermediate ADaM: a use case

Definition: Time to increased glucose level



From the SAP:

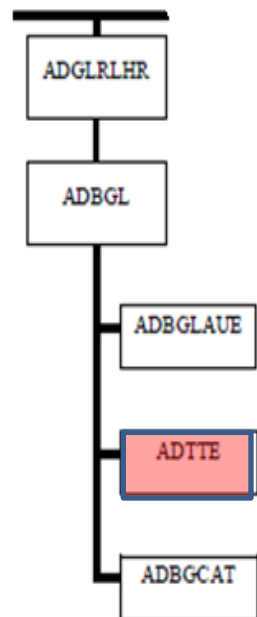
Time to the earliest significantly increased blood glucose level in hours will be defined as the time from administration of study treatment to the first blood glucose level that is ≥ 2 SDs above the 72-hour baseline average blood glucose level (relative hour of first increased blood glucose level). This will be done for the following time periods:

- Hour 1-72
- Hour 1-360

Intermediate ADaM: a use case

Step 3 : Windowing time-point and visit

ADTTE – Time to Event AD



- Time to Increased Blood Glucose
- Derived from ADBGL

PARAM	AVAL	CNSR	EVNTDESC
Time to Increased Blood Glucose Levels (Hours) 1-72	72	1	NO EVENT
Time to Increased Blood Glucose Levels (Hours) 1-360	180	0	>= BASELINE MEAN + 2SD

SRCDOM	SRCVAR	SRCSEQ
ADBGL	ATPTN	143
ADBGL	ATPTN	380

Data-Point
Traceability to
Intermediate ADaM

Conclusions

The “Art” of being traceable

- Traceability **the art** of making complex things simple, clear and transparent
- **Splitting** complex processes into smaller pieces facilitates the **understanding** and the **reproducibility** (validation)
- Do not forget about being **analysis-ready**, one of the four ADaM fundamental principles, this is also about traceability

Backup Slides

Standard	Data Point Traceability	Metadata Traceability (Supportive Documents)
SDTM	«Not applicable/Not needed»	aCRF define.xml SDRG
ADaM	Copy SDTM variables --SEQ from SDTM SRCDOM/SRCVAR/SRCSEQ ADTF ASEQ DTYPE ANLxxFL Occurrence Flags in OCCDS Intermediate ADaMs	define.xml ADRG SAP
Analysis Results	«Not applicable/Not needed»	define.xml (ARM extension) ADRG SAP