

Business & Decision

Life Sciences

The value of optimising
your CDISC Analysis Data
Model (ADaM)
specifications

Amine Khemiri

DS06 – PHUSE Virtual EU Connect 2021



Business & Decision AGENDA Life Sciences

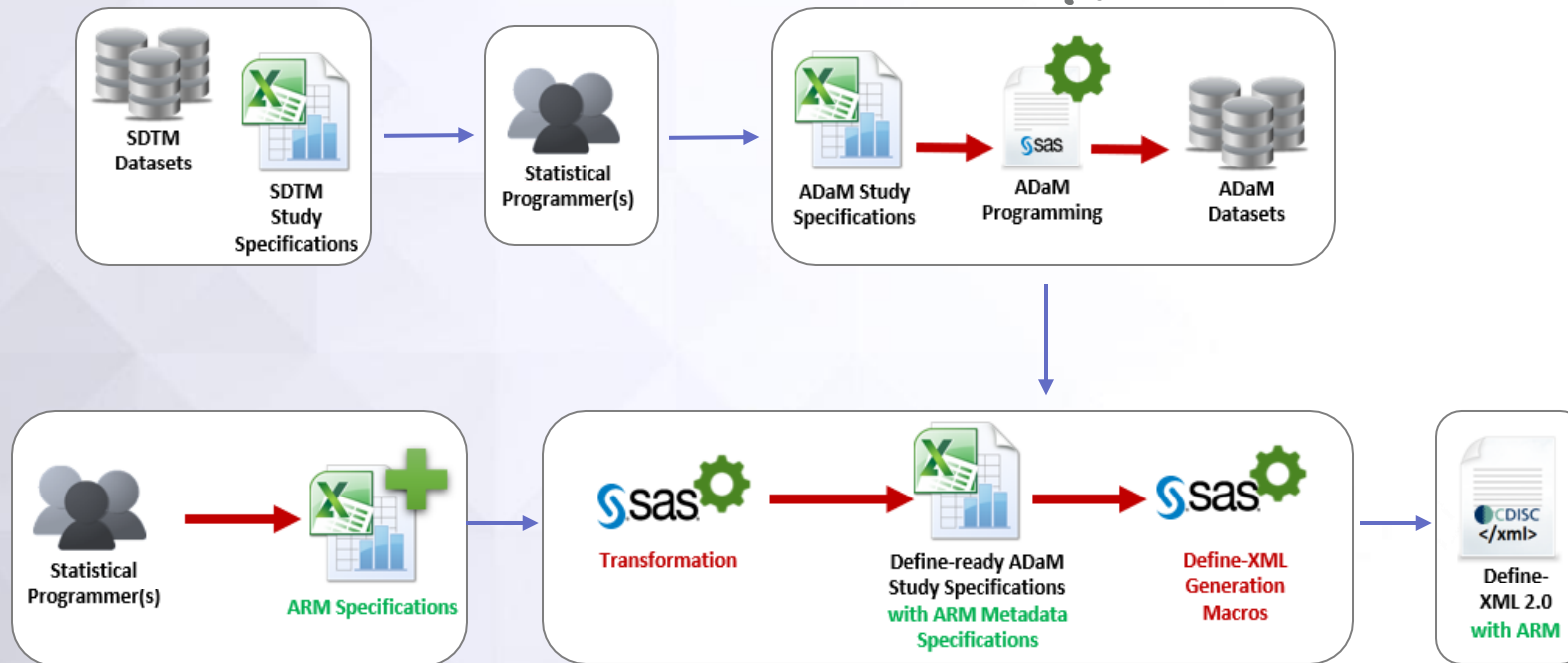
- Background
- Issues & Challenges
- Findings & Proposed Solution
- Best Practices
- Summary & Conclusion



#Data #DataDriven
#DataStrategy #DataIntegration
#DataViz #DataStandards
#DataScience #Datagovernance
#HealthCare #DataAnalysis
#ClinicalTrials
#ArtificialIntelligence
#AdvancedAnalytics
#MachineLearning #Digital
#Chatbot #Biometrics #CRO
#ClinicalData
#LifeSciencesExperts
#Whodrug #MedDRA
#Pharmacovigilance
#ClinicalResearchLifeCycle
#MedicalWriting
#MedicalInformation
#MedicalReview #SDTM
#ADaM #CDISC #Innovation
#RegulatorySubmissions
#Protocol #Coding

Business & Decision BACKGROUND

Life Sciences



- Inputs:
 - Study protocol, SAP
 - aCRF, Shells
 - SDTM datasets
- Outputs:
 - ADaM datasets
 - TFLs
 - Define-XML
 - ADRG

Business & Decision

ADaM SPECIFICATIONS (MAPPING)

Life Sciences

- Detail how to create an ADaM dataset:

- How it is structured

- ADSL / BDS / OCCDS
- Other

Dataset	Description	Structure	Purpose	Class
ADSL	Subject-Level Analysis Dataset	One record per subject	Analysis	ADSL
ADSV	Subject Visits Analysis Dataset	One record per subject per analysis visit	Analysis	OTHER
ADMH	Medical History Analysis Dataset	One record per subject per medical history per date	Analysis	OCCDS
ADFA	Findings Analysis Dataset	One record per subject per parameter per analysis visit	Analysis	BDS
ADVS	Vital Signs Analysis Dataset	One record per subject per parameter per analysis visit	Analysis	BDS
ADPE	Physical Examination Analysis Dataset	One record per subject per parameter per analysis visit	Analysis	BDS
ADAE	Adverse Events Analysis Dataset	one record per subject per adverse event	Analysis	OCCDS

- How these variables are created

- Predecessor
- Derived

- What variables you need

DOMAIN PREFIX	VARIABLE NAME	VARIABLE LABEL	VARIABLE TYPE	VARIABLE LENGTH	CDISC CT	COMP. ALGORITHM	ORIGIN (name of ADaM dataset or derived)
ADSL	STUDYID	Study Identifier	char	10			DM
ADSL	USUBJID	Unique Subject Identifier	char	18			DM
ADSL	SUBJID	Subject Identifier for the Study	char	7			DM
ADSL	SITEID	Study Site Identifier	char	4			DM
ADSL	ARM	Description of Planned Arm	char	10	ARM		DM
ADSL	ARMN	Description of Planned Arm (N)	num	8	ARMN	ADSL.ARMN	Derived
ADSL	ARMCD	Planned Arm Code	char	1	ARMCD		DM
ADSL	ACTARM	Description of Actual Arm	char	10	ARM		DM
ADSL	ACTARMN	Description of Actual Arm (N)	num	8	ARMN	ADSL.ACTARMN	Derived

Business & Decision

ADaM COMPLIANCE

Life Sciences

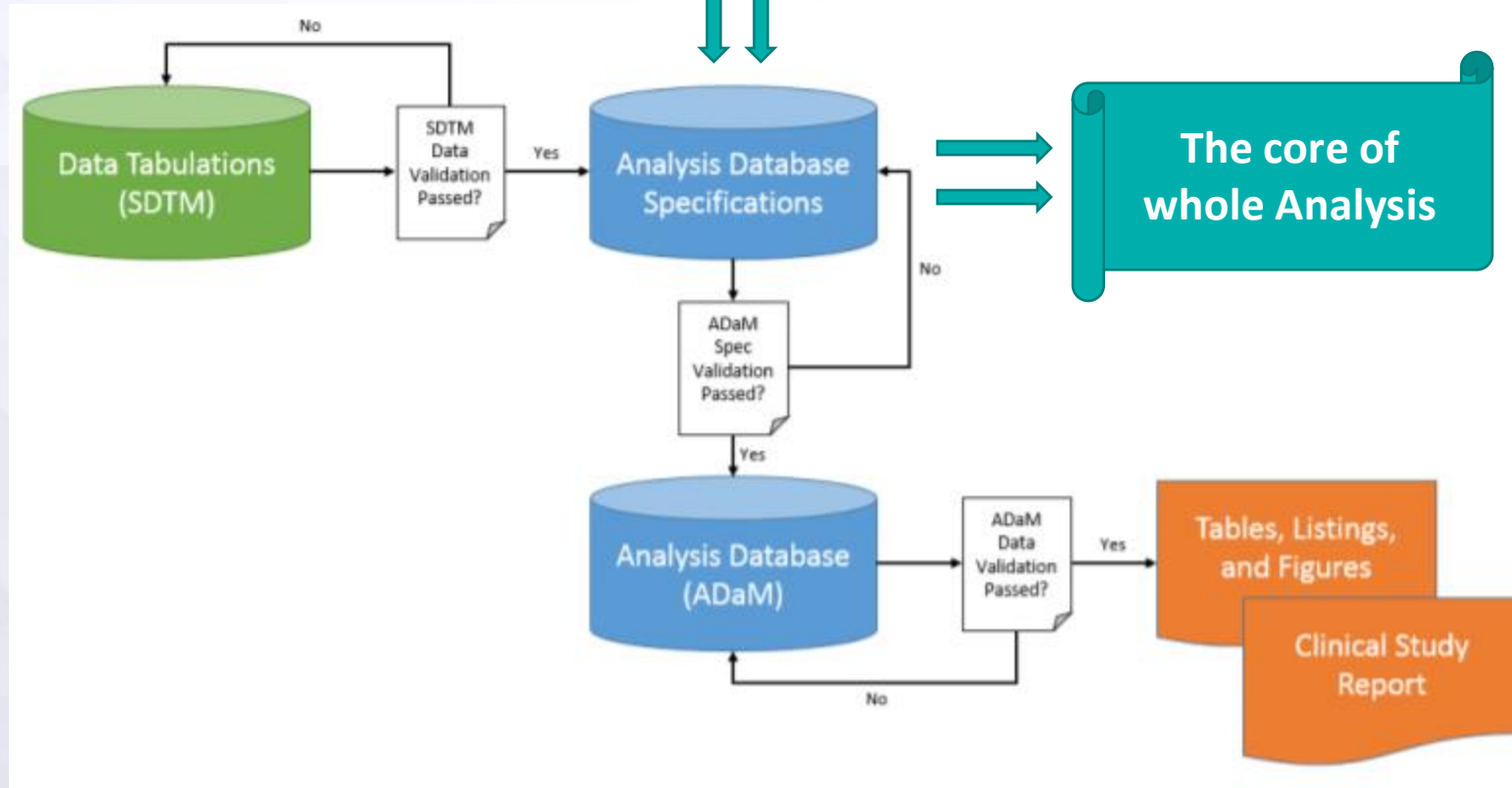
- ADaM Compliance Starts with ADaM Specifications
 - Follow the fundamental principles described in the ADaM model document
 - Follow CDISC ADaM Implementation Guide ADaM-IG v 1.2
- Robust, clean, and CDISC compliant specifications:
 - Constantly evolving throughout a study
 - A lot of checks
 - Timelines

How can we reduce development cost?

Business & Decision WORKFLOW

Life Sciences

Optimization



- The list of optimisation techniques includes but is not limited to:

Variables needed
for analysis

Save SDTM findings

Follow BDS/OCCDS
structure

Not copying all
variables from ADSL
to other ADaMs

Length of character
variables

Plan tasks in
advance

Log questions and
decisions

Business & Decision

BEST PRACTICES

Life Sciences

1. Do not copy all variables from ADSL to other ADaM datasets

- Why not?

- More risk of errors (quality)
- Avoid re-doing the work (time)

- How?

- Copy only relevant variables
- Keep the principle “Analysis-ready”

DOMAIN PREFIX	VARIABLE NAME	VARIABLE LABEL	VARIABLE TYPE	VARIABLE LENGTH	CDISC CT	COMP. ALGORITHM	ORIGIN (name of ADAM dataset or derived)
ADSL	STD1PER	Discontinue the Study <=3 weeks?	char	1	Y BLANK	ADSL.STD1PER	Derived
ADSL	STD2PER	Discontinue the Study <=6 weeks?	char	1	Y BLANK	ADSL.STD2PER	Derived
ADSL	STD3PER	Discontinue the Study <=9 weeks?	char	1	Y BLANK	ADSL.STD3PER	Derived
ADSL	STD4PER	Discontinue the Study <=12 weeks?	char	1	Y BLANK	ADSL.STD4PER	Derived
ADSL	STD5PER	Discontinue the Study <=15 weeks?	char	1	Y BLANK	ADSL.STD5PER	Derived
ADSL	STD6PER	Discontinue the Study <=18 weeks?	char	1	Y BLANK	ADSL.STD6PER	Derived
ADSL	STD7PER	Discontinue the Study <=21 weeks?	char	1	Y BLANK	ADSL.STD7PER	Derived
ADSL	STD8PER	Discontinue the Study <=24 weeks?	char	1	Y BLANK	ADSL.STD8PER	Derived

DOMAIN PREFIX	VARIABLE NAME	VARIABLE LABEL	VARIABLE TYPE	VARIABLE LENGTH	CDISC CT	COMP. ALGORITHM	ORIGIN (name of ADAM dataset or derived)
ADAE	STD1PER	Discontinue the Study <=3 weeks?	char	1	Y BLANK		ADSL
ADAE	STD2PER	Discontinue the Study <=6 weeks?	char	1	Y BLANK		ADSL
ADAE	STD3PER	Discontinue the Study <=9 weeks?	char	1	Y BLANK		ADSL
ADAE	STD4PER	Discontinue the Study <=12 weeks?	char	1	Y BLANK		ADSL
ADAE	STD5PER	Discontinue the Study <=15 weeks?	char	1	Y BLANK		ADSL
ADAE	STD6PER	Discontinue the Study <=18 weeks?	char	1	Y BLANK		ADSL
ADAE	STD7PER	Discontinue the Study <=21 weeks?	char	1	Y BLANK		ADSL
ADAE	STD8PER	Discontinue the Study <=24 weeks?	char	1	Y BLANK		ADSL

2. Follow BDS/OCCDS structure as much as possible

- Why?
 - Reviewers are more familiar with BDS/OCCDS
 - Follow Implementation Guide when designing
 - Avoid explanation of the “Other” structure in the ADaM Reviewer’s Guide

Dataset	Description	Structure	Purpose	Class
ADSL	Subject-Level Analysis Dataset	One record per subject	Analysis	ADSL
ADSV	Subject Visits Analysis Dataset	One record per subject per analysis visit	Analysis	OTHER
ADMH	Medical History Analysis Dataset	One record per subject per medical history per date	Analysis	OCCDS
ADFA	Findings Analysis Dataset	One record per subject per parameter per analysis visit	Analysis	BDS
ADVS	Vital Signs Analysis Dataset	One record per subject per parameter per analysis visit	Analysis	BDS
ADPE	Physical Examination Analysis Dataset	One record per subject per parameter per analysis visit	Analysis	BDS
ADAE	Adverse Events Analysis Dataset	one record per subject per adverse event	Analysis	OCCDS

Business & Decision

BEST PRACTICES

Life Sciences

3. Set length of character variables to maximum when you start

- Why?

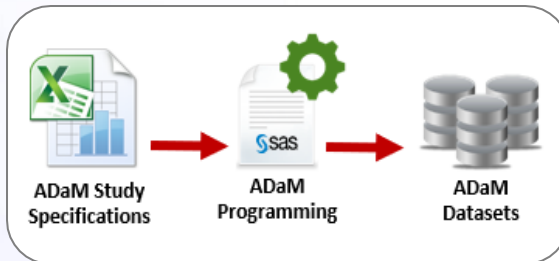
- Avoid truncation issues (quality)
- Do not review and fix at each release (time)

DOMAIN PREFIX	VARIABLE NAME	VARIABLE LABEL	VARIABLE TYPE	VARIABLE LENGTH
ADCM	STUDYID	Study Identifier	Char	200
ADCM	USUBJID	Unique Subject Identifier	Char	200
ADCM	SUBJID	Subject Identifier for the Study	char	200
ADCM	SITEID	Study Site Identifier	char	200
ADCM	CMTRT	Reported Name of Drug, Med, or Ther	Char	200
ADCM	CMDECOD	Standardized Medication Name	Char	200
ADCM	CMCAT	Category for Medication	Char	200

Ongoing updates

DOMAIN PREFIX	VARIABLE NAME	VARIABLE LABEL	VARIABLE TYPE	VARIABLE LENGTH
ADCM	STUDYID	Study Identifier	Char	10
ADCM	USUBJID	Unique Subject Identifier	Char	18
ADCM	SUBJID	Subject Identifier for the Study	char	7
ADCM	SITEID	Study Site Identifier	char	4
ADCM	CMTRT	Reported Name of Drug, Med, or Ther	Char	120
ADCM	CMDECOD	Standardized Medication Name	Char	76
ADCM	CMCAT	Category for Medication	Char	39

Final release



Business & Decision

BEST PRACTICES

Life Sciences

4. Plan tasks in advance

- Why?
 - To check what will be relevant
 - To avoid developing heavy specifications
- How?

Prepare questions



Track changes

DOMAIN PREFIX	VARIABLE NAME	VARIABLE LABEL	VARIABLE TYPE	VARIABLE LENGTH	CDISC CT	COMP. ALGORITHM
ADSL	STUDYID	Study Identifier	char	10		
ADSL	USUBJID	Unique Subject Identifier	char	18		
ADSL	SUBJID	Subject Identifier for the Study	char	7		
ADSL	SITEID	Study Site Identifier	char	4		
ADSL	ITTFL	Intent-To-Treat Population Flag	char	1	YN	ADSL.ITTFL
ADSL	FASFL	Full Analysis Set Population Flag	char	1	YN	ADSL.FASFL
ADSL	SAFFL	Safety Population Flag	char	1	YN	ADSL.SAFFL
ADSL	PPROTFL	Per-Protocol Population Flag	char	1	YN	ADSL.PPROTFL

5. Log questions and decisions in a specific document

Detailed description	Corrected by?	Date Corrected	Resolution and Action taken (includes mapping decisions and date reported to the sponsor)	Is the issue resolved ?
Can we add reproductive information (use RP domain see p. 5 aCRF)	AKH	11-Mar-21	I have added reproductive information to ADSL.	Yes
Add ISO8601 to column J, where appropriate. Eg. For variables RFSTDTC and RFENDTC	AKH	11-Mar-21	Done.	Yes
Shouldn't we add the variables of aCRF p8-11 (cancer staging/mutational diagnosis/... variables)? For example: STAGED, Following the SAP: In the demography listing, if a subject has more than one race recorded, then all races selected will be listed for that subject. Should we also list RACEOTH (p4 aCRF). Please add to ADSL if needed for listing.	AKH	11-Mar-21	They will be added in ADFA.	Yes
https://www.lexjansen.com/pharmasug/2018/SS/PharmaSUG-2018-SS15.pdf variables coming from QVAL, QNAM, will these be seen as derived or copied?	AKH	11-Mar-21	I will check if it is possible to make this "Predecessor " when generating the define.	No
Race has a fixed codelist, which doesn't include 'OTHER' and 'MULTIPLE'. This might lead to some warning or error in p21, as we have 'OTHER' and 'MULTIPLE' as option in the codelist now. Let's check this when we can generate define/p21.	AKH	12-Mar-21	To be checked when we generate the define.	No

6. Save SDTM findings in a dedicated file.

Detailed description	Corrected by?	Date Corrected	Resolution and Action taken (includes mapping decisions and date reported to the sponsor)	Is the issue resolved ?
I think RFENDTC and RFXENDTC should be in date/time format in order to be consistent with RFSTDTC and RFXSTDTC. Could you please have a look onto this?	AKH	16-Mar-21	To keep as is. RFENDTC and RFXENDTC will contain the date of last dose with the treatment which is captured in the DA domain where not time is collected. Not that the RFXENDTC is also expected to be the same at the max EXENDTC.	Yes
I think RFPENDTC is a mix between datetime and date formats. Could you please check if end time is not missing in raw data?	AKH	16-Mar-21	I don't know where the time part is coming from for this variable actually. Time part should not be part of this variable. Also when EOS date is not completed yet RFPENDTC is left blank while latest visit date can be taken.	No
I think EXENDTC should be in date/time format in order to be consistent with EXSTDTC. Could you please have a look onto this?	AKH	16-Mar-21	To keep as is. EXENDTC will collect the date of last dose which is retrieved from the DA domain. Only date and no time is collected.	Yes
Visitnum is not populated	AKH	7-Jun-21	VISITNUM can not be populated as information not collected in eCRF. The field will stay empty in SDTM	QC OK
There are queries on the EC form for both visits. They indicated to start the new kit on the day of the visit before it was even assigned which is not possible. To update dates on DA form and update the kit number assigned on the EC form	PYO	7-Jun-21	Source data issues: To check with DM if all queries posted in EDC. Do you have a list of identified cases?	QC OK
I think RACE codelists "OTHER" "MULTIPLE" are not part of CDISC Controlled Terminology. Please update.	PYO	7-Jun-21	Values OTHER and MULTIPLE added to cover the fact that investigator can tick several options in eCRF (in accordance with FDA CDASH recommendations). <u>Explanation will be reported in Reviewer Guide (RG)</u>	QC OK

7. Select only variables needed for analysis and traceability

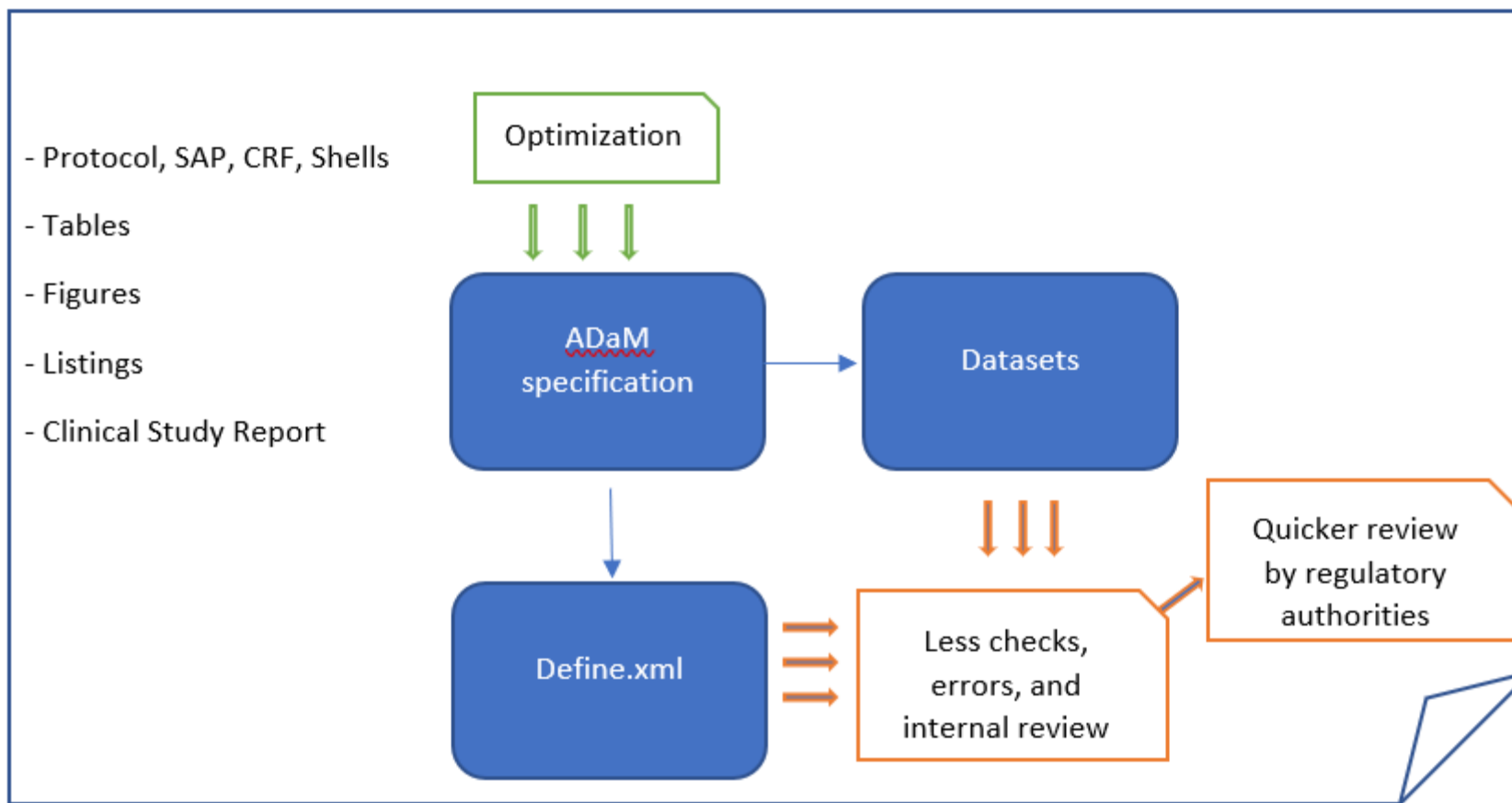
DOMAIN PREFIX	VARIABLE NAME	VARIABLE LABEL	VARIABLE TYPE	VARIABLE LENGTH	CDISC CT	COMP. ALGORITHM	ORIGIN (name of ADAM dataset or derived)
ADSL	STUDYID	Study Identifier	char	10			DM
ADSL	USUBJID	Unique Subject Identifier	char	18			DM
ADSL	SUBJID	Subject Identifier for the Study	char	7			DM
ADSL	SITEID	Study Site Identifier	char	4			DM
ADSI	ARM	Description of Planned Arm	char	10	ARM		DM
ADSL	ITTFL	Intent-To-Treat Population Flag	char	1	YN	ADSL.ITTFL	Derived
ADSL	FASFL	Full Analysis Set Population Flag	char	1	YN	ADSL.FASFL	Derived
ADSL	SAFFL	Safety Population Flag	char	1	YN	ADSL.SAFFL	Derived
ADSL	PPROFL	Per-Protocol Population Flag	char	1	YN	ADSL.PPROFL	Derived
ADSL	HEIGHTBL	Baseline Height (cm)	num	8		ADSL.HEIGHTBL	Derived
ADSL	WEIGHTBL	Baseline Weight (kg)	num	8		ADSL.WEIGHTBL	Derived
ADSL	BMIBL	Baseline BMI (kg/m ²)	num	8		ADSL.BMIBL	Derived

Table 14.1.3.1.1. Demographic Characteristics - Intent-to-Treat Set			
	Drug A (N=XXX)	Drug B (N=XXX)	Overall (N=XXX)
Gender [n (%)]			
Male	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Female	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
Age at baseline (years)			
n	xxx	xxx	xxx
Mean (SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
Median	xx.x	xx.x	xx.x
Q1, Q3	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
Min, Max	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
Height (cm)			
n	xxx	xxx	xxx
Mean (SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
Median	xx.x	xx.x	xx.x
Q1, Q3	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
Min, Max	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
Weight (kg)			
n	xxx	xxx	xxx
Mean (SD)	xx.x (xx.x)	xx.x (xx.x)	xx.x (xx.x)
Median	xx.x	xx.x	xx.x
Q1, Q3	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
Min, Max	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x

Business & Decision

SUMMARY AND CONCLUSION

Life Sciences





THANK YOU FOR YOUR ATTENTION
Q&A

Business & Decision CONTACT Life Sciences



Caroline Pelckmans

Statistical Programmer/ Biostatistician

Caroline.pelckmans@businessdecision.be



Amine Khemiri

Statistical Programmer

amine.khemiri@businessdecision.be



Pamphile Beaujean

Statistical Programmer

pamphile.beaujean@businessdecision.be