cdisc

2023
EUROPE
INTERCHANGE
COPENHAGEN | 26-27 APRIL

Balancing quality and efficiency - a Novo Nordisk submission example

Presented by Adrian Czaban, Statistical Programming Specialist,
Biostatistics 2, Novo Nordisk

Meet the Speaker

Adrian Czaban

Title: Statistical Programming Specialist/International Lead Programmer

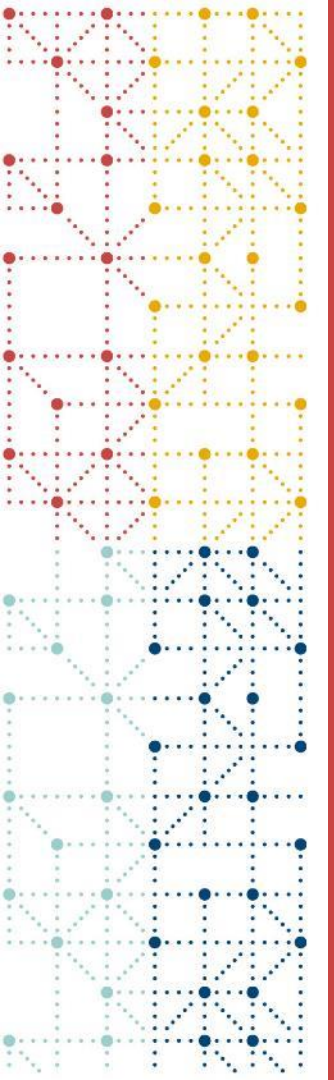
Organization: Biostatistics 2, Novo Nordisk

- Started programming career in 2015
- Happy to work on new things and try out new solutions
- Lives with wife and 2 daughters in Copenhagen area

Disclaimer and Disclosures

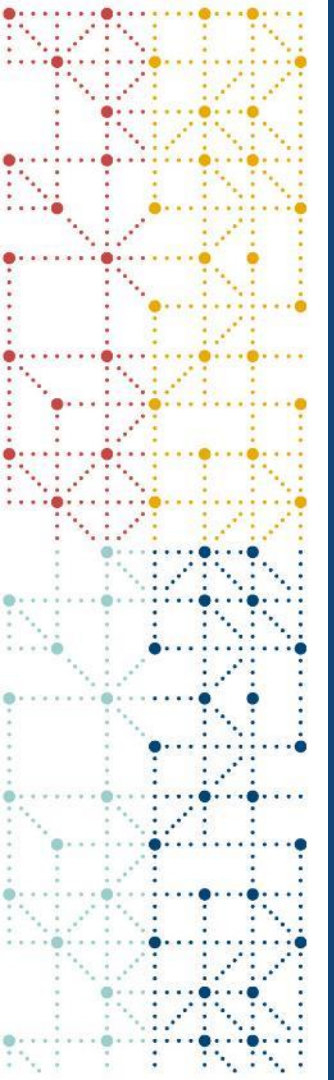
- *The views and opinions expressed in this presentation are those of the author and do not necessarily reflect the official policy or position of CDISC.*

- *The author has no real or apparent conflicts of interest to report.*



Agenda

1. Introduction to the project
2. Programming setup
3. Opportunities and challenges
4. Outcome



Introduction to the project and deliverables

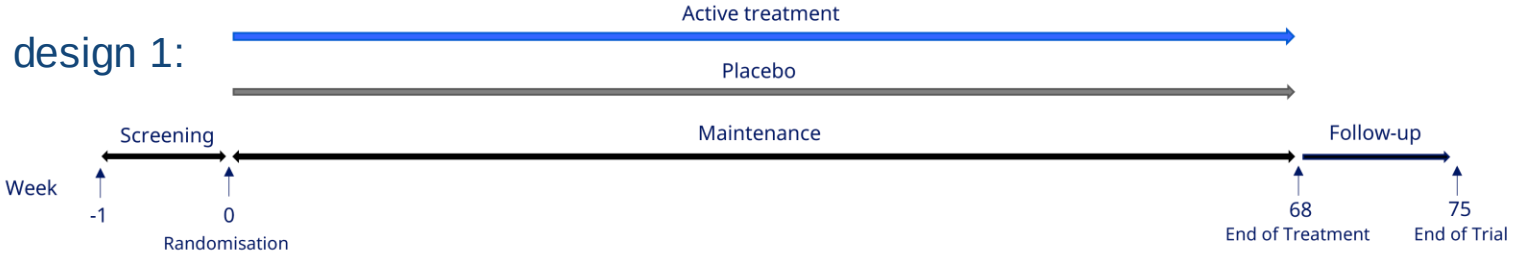
Large project with the goal of submitting for a NDA in all major markets

4 phase 1 trials

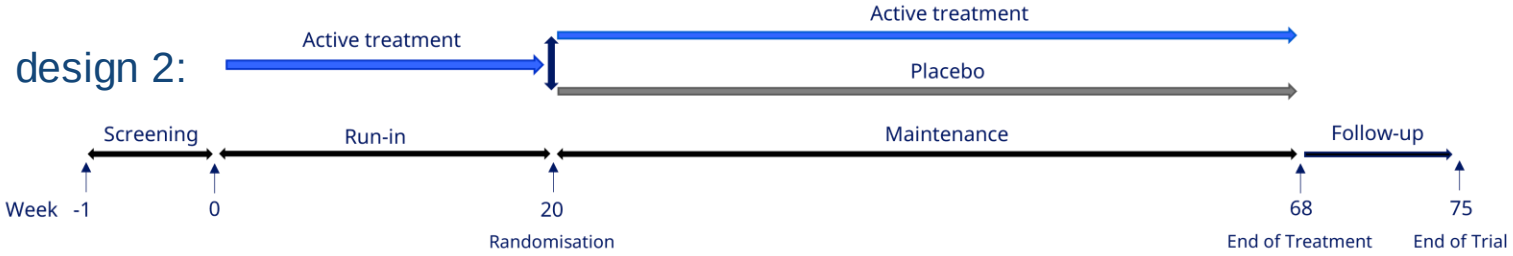
1 phase 2 trial

5 phase 3a trials with 2 types of design

Trial design 1:



Trial design 2:



Contents of the submission package

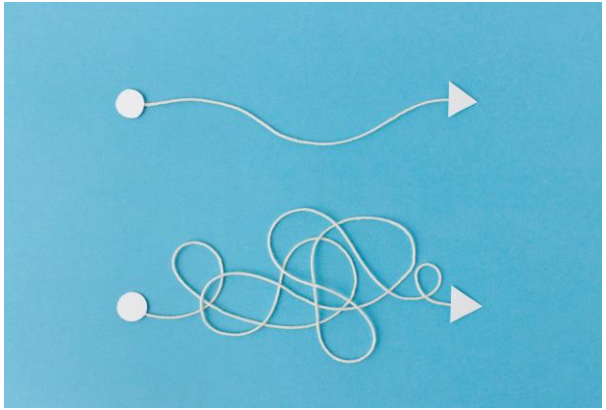
- For each trial and summaries, we planned to include the following:
  - ADRG
  - define.xml with Analysis Results Metadata (ARM)
  - ADaM datasets
  - All programs creating ADaM datasets and all macros used
  - All metadata datasets along with their programs, apart from programming plan specification
  - A subset of output programs where we have defined the ARM. This was split into programs creating the final output as well as programs making the statistical analysis
  - BIMO

What would we like to achieve?

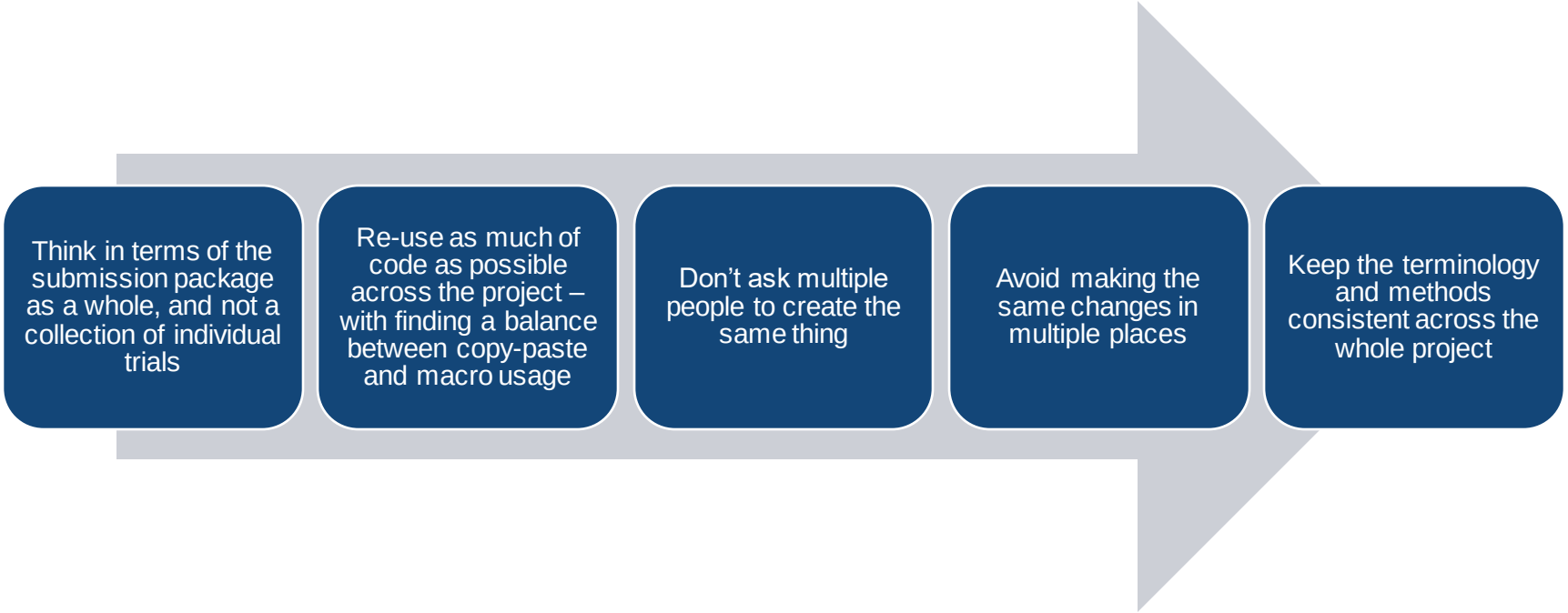
Programs that are ”submission ready”: easy to read and understand, and can be easily executed by FDA if required. Consistent both internally and across the project.

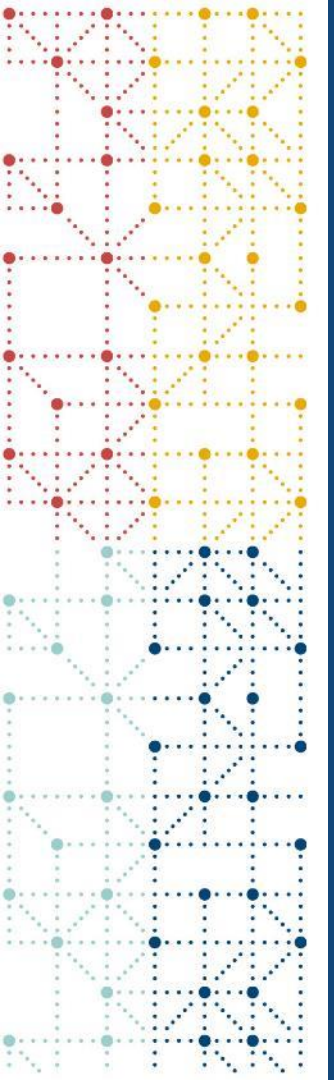
ADaM datasets, define.xml and and ADRG deliverables conforming with CDISC requirements and in tight agreement with each other.

Make the review process as smooth as possible, by keeping things simple and giving reviewers the tools to find anwsers themselves.



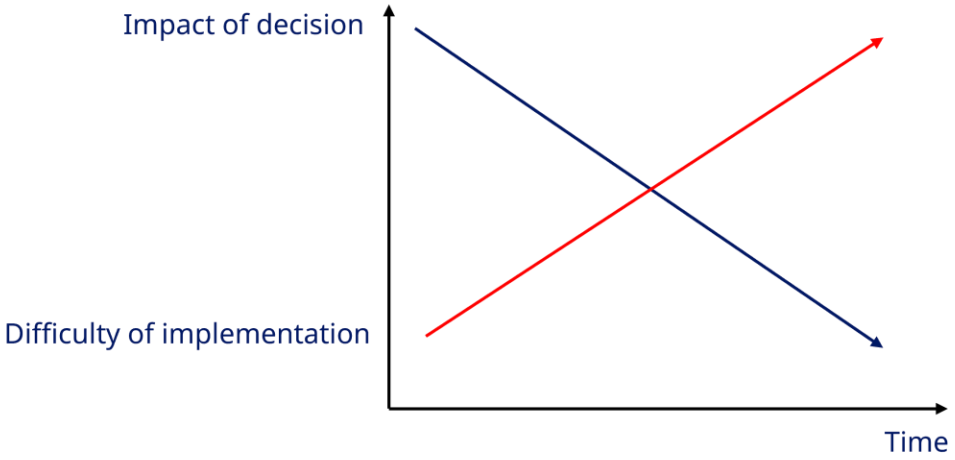
What can we do to be more efficient when working in our trials – main focus points





Programming setup

Aligning already on the CRF design stage



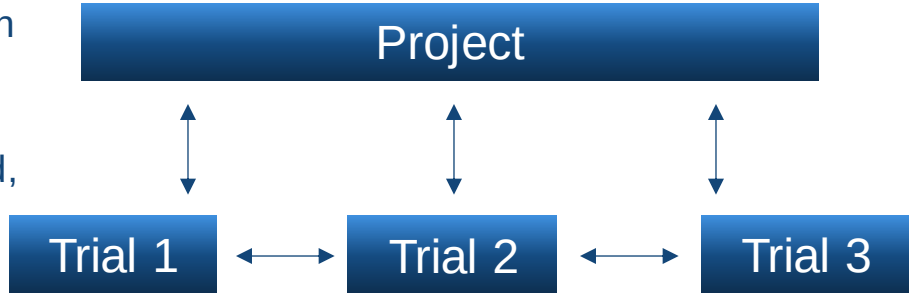
Same forms

Same definitions

Data issues relevant to the entire project shared across all people.

Alignment and ease of combining the trials

- All ADaM and output programs first created for Trial nr 1
- All of the programmers assigned their own ADaM programs, that are shared in a common project space
- Once Trial 1 programs have been created, programmers have moved out into their respective trials
- Similarly, output programs have been designed to work across all trials and maintained in the project space



How do we keep a common dataset content?

Project data content sheet, ensuring the same content definitions are used across all trials

include_in_Trial1	include_in_Trial2	include_in_Trial3	include_in_Trial4	include_in_Trial5	include_in_pooled_database	Table	Column	Label
Y	Y	Y	Y	Y	Y	ADSL	ARMCD	Planned Arm Code
Y	Y	Y	Y	Y	Y	ADSL	ACTARM	Description of Actual Arm
Y	Y	Y	Y	Y	Y	ADSL	ACTARMCD	Actual Arm Code
Y	N	Y	Y	N	N	ADSL	ACTARMN	Numeric version of Actual Arm (N)
N	Y	N	N	N	Y	ADSL	ACTARMN	Numeric version of Actual Arm (N)
Y	Y	Y	Y	Y	Y	ADSL	FASFL	Full Analysis Set Population Flag
Y	Y	Y	Y	Y	Y	ADSL	FASREA	Reason for Exclusion from FAS
Y	Y	Y	Y	Y	Y	ADSL	SAFFL	Safety Population Flag
Y	Y	Y	Y	Y	Y	ADSL	SAFREA	Reason for Exclusion from SAF
Y	Y	Y	Y	Y	Y	ADSL	PPROTFL	Per-Protocol Population Flag

algorithm	algorithm_Trial1	algorithm_Trial2	algorithm_Trial3	algorithm_Trial4	algorithm_Trial5	algorithm_pool
When ANL04FL eq Y				When ANL04FL eq Y		When ANL04FL eq Y

Thinking of project instead of trials – pooled database developed alongside other tasks

```
1  /*************************************************************************
2  Description: Development of Laboratory Assessments(ADLB) dataset of 7  trials.
```

<Stack datasets>
<Derive pool-specific variables>

```
156 ************************************************************************;
157 **** STEP 6: OUTPUT POOLED ADLB                                       ****;
158 ************************************************************************;
159 %adamensurestruct(
160    inds                        = adlb
161    , remove_non_specified_cols = Y
162    , missing_column_severity   = WARN
163    );
```

How do we align on output level?

Trial 1 : 351

Trial 2 : 371

Trial 3 : 315

Trial 4 : 322

Trial 5 : 383

Summaries : 622

Total: 2364

What could go wrong?

Age (years)
N
18-<65
65-<75
75-<85
>=85

Age (years)
N
18-65
65-<75
75-<85
>=85

Age (years)
N
>=85
75-<85
65-<75
18-<65

Age (years)
N
18-<75
75-<85
>=85

Strict control of codelists on the project level

Master file:

include_in_trial_1	include_in_trial_2	codelist	value	ord	descrpt
Y	Y	agegrp1skl	N	5	N
Y	Y	agegrp1skl	18-<65	10	18-<65
Y	Y	agegrp1skl	65-<75	20	65-<75
Y	Y	agegrp1skl	75-<85	30	75-<85
Y	Y	agegrp1skl	>=85	40	>=85

Metadata:

CODELIST	CODE	ORDER	DECODE
agegrp1skl	N	5	N
agegrp1skl	18-<65	10	18-<65
agegrp1skl	65-<75	20	65-<75
agegrp1skl	75-<85	30	75-<85
agegrp1skl	>=85	40	>=85

Output:

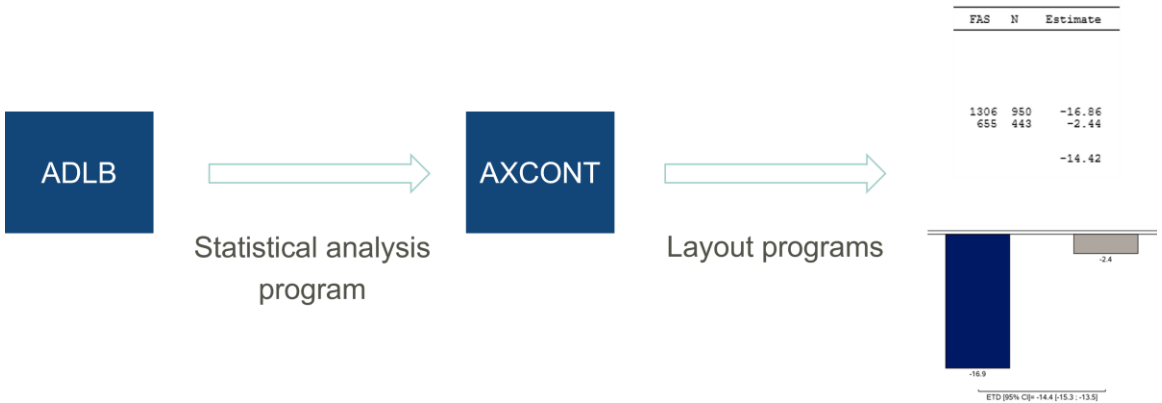
```
Age (years)
  N
  18-<65
  65-<75
  75-<85
  >=85
```

Usage of macros

Previous feedback on too many/too complicated macro programs

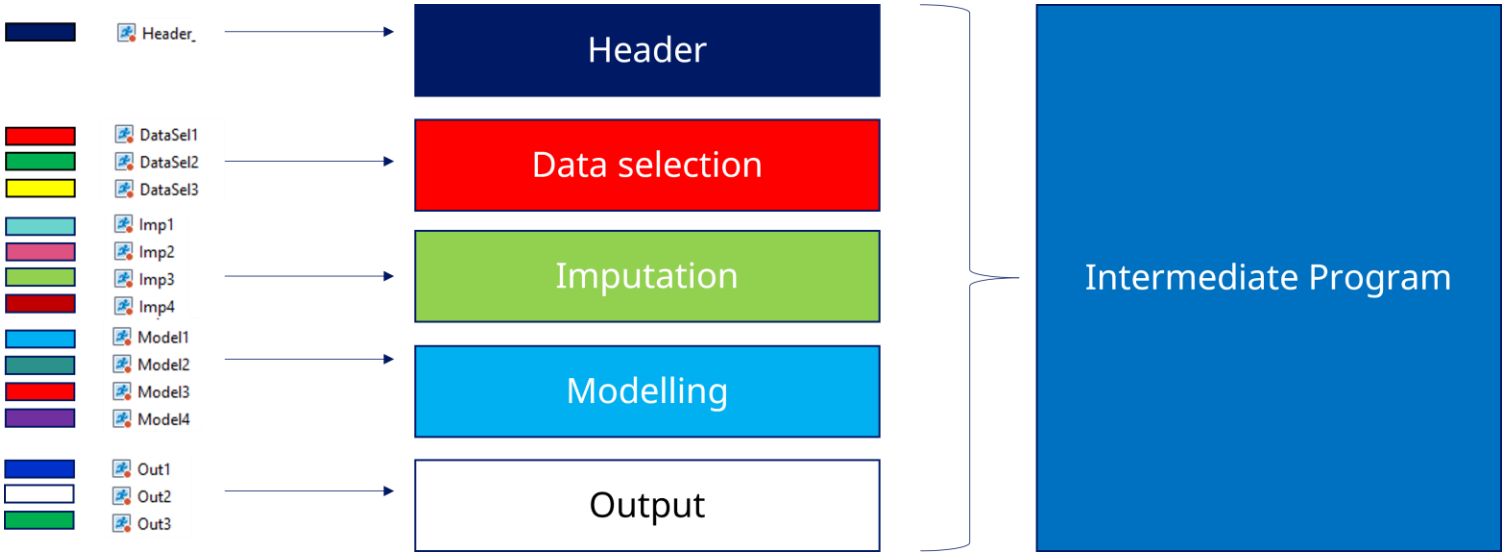
Statistical analysis without macros – ‘AX programs’ setup

- Programs making statistical analyses are some of the most complex to review and understand.
- However, there is a danger of creating complex macro programs which are hard to debug and take a very long time to execute.
- Introducing an intermediate step and separating the program making the analyses from the one generating the output has many benefits



Generating programs: 1 analysis – 1 program approach

- All AX programs should be self-contained and submission-ready
- There is a need to generate those programs in a simple and automated way
- Solution – building blocks



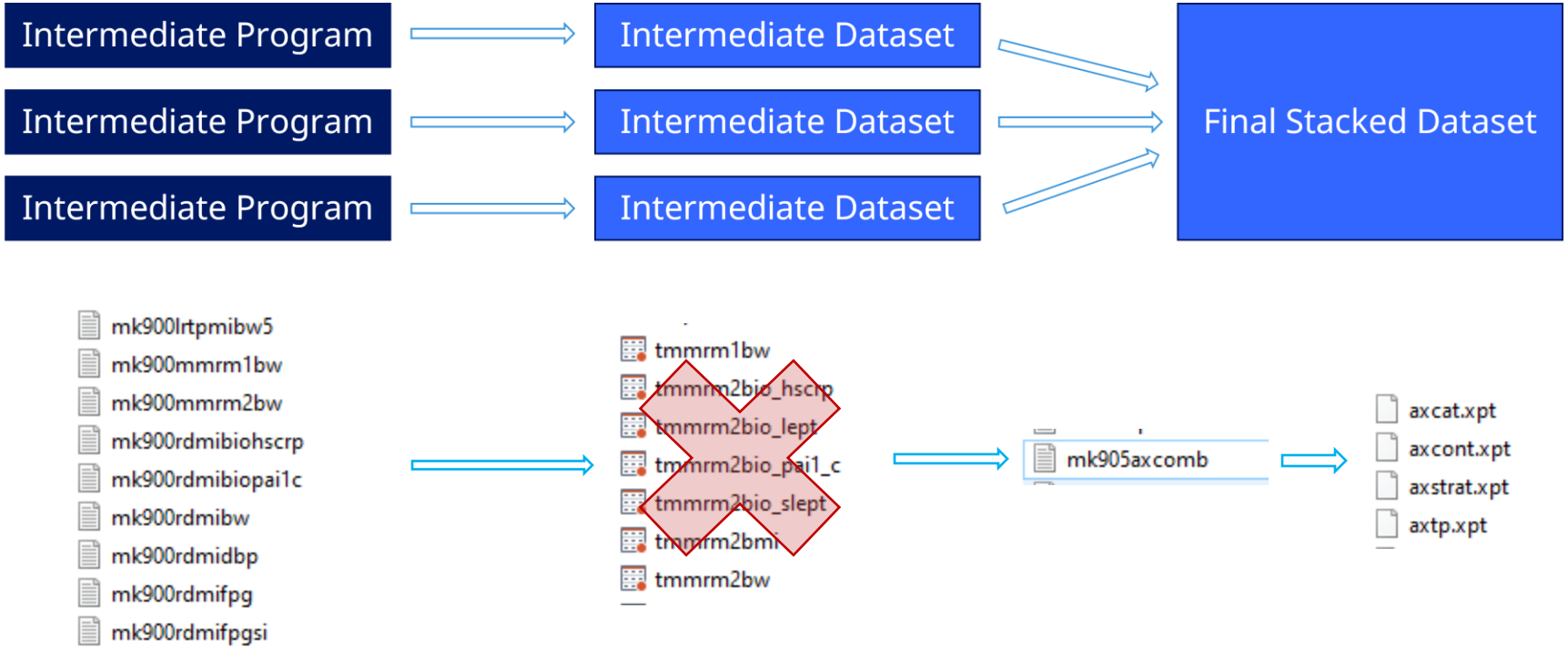
Two levels of control:

- Which blocks to use

		Header				Data selection				Imputation				Modelling				Output			
		HeaderX				DataSelX				ImpX				ModelX				OutX			
Model/imputation type	Program part-name	TRIAL 1	TRIAL 2	TRIAL 3	TRIAL 4	TRIAL 1	TRIAL 2	TRIAL 3	TRIAL 4	TRIAL 1	TRIAL 2	TRIAL 3	TRIAL 4	TRIAL 1	TRIAL 2	TRIAL 3	TRIAL 4	TRIAL 1	TRIAL 2	TRIAL 3	TRIAL 4
ANCOVA - RD-MI	rdmi	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
LR - RD-MI	lrrdmi	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
ANCOVA - J2R-MI	j2rmi	1	1	1	1	1	1	1	1	2	2	2	2	1	1	1	1	1	1	1	1
LR - J2R-MI	lrj2rmi	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2
ANCOVA - S1-SI	s1	1	1	1	1	1	1	1	1	3	3	3	3	3	3	3	3	1	1	1	1
LR - S1-SI	lrs1	1	1	1	1	1	1	1	1	3	3	3	3	4	4	4	4	2	2	2	2
ANCOVA - S2-SI	s2	1	1	1	1	1	1	1	1	3	3	3	3	3	3	3	3	1	1	1	1
LR - S2-SI	lrs2	1	1	1	1	1	1	1	1	3	3	3	3	4	4	4	4	2	2	2	2
ANCOVA - TP-MI	tpmi	1	1	1	1	1	1	1	1	1	1	1	1	5	5	5	5	3	3	3	3
LR - TP-MI	lrtpmi	1	1	1	1	1	1	1	1	1	1	1	1	6	6	6	6	4	4	4	4
MMRM1	mmrm1	1	1	1	1	1	1	1	1	NA	NA	NA	NA	7	7	7	7	1	1	1	1
LR - MMRM1	lrmmrm1	1	1	1	1	1	1	1	1	4	4	4	4	4	4	4	4	2	2	2	2
MMRM2	mmrm2	1	1	1	1	1	1	1	1	NA	NA	NA	NA	7	7	7	7	1	1	1	1
LR - MMRM2	lrmmrm2	1	1	1	1	1	1	1	1	4	4	4	4	4	4	4	4	2	2	2	2
LR - NR	lrnr	1	1	1	1	1	1	1	1	5	5	5	5	4	4	4	4	2	2	2	2
LR - RD-MI - CO	lrrdmico		1				2				6				8				5		
LR - MMRM2 - CO	lrmmrm2co		1				2				7				9				5		

- Make slight changes to individual blocks

model	estimand	param1	param2	dataset1	dataset2	chgtype1	chgtype2	impfactors
$50.	$30.	$200.	$200.	$9.	$9.	$12.	$12.	$50.
ANCOVA - RD-MI	Treatment policy	Body Weight (kg)		ADVS		PCHG		SEX BMIG3BL STRAT1V STRAT2V
LR - RD-MI	Treatment policy	Body Weight (kg)		ADVS		PCHG		SEX BMIG3BL STRAT1V STRAT2V
LR - RD-MI	Treatment policy	Body Weight (kg)		ADVS		PCHG		SEX BMIG3BL STRAT1V STRAT2V
LR - RD-MI	Treatment policy	Body Weight (kg)		ADVS		PCHG		SEX BMIG3BL STRAT1V STRAT2V
ANCOVA - RD-MI	Treatment policy	Waist Circumference (cm)		ADVS		CHG		SEX BMIG3BL STRAT1V STRAT2V
ANCOVA - RD-MI	Treatment policy	Body Weight (kg)		ADVS		PCHG		SEX BMIG3BL STRAT1V STRAT2V
ANCOVA - RD-MI	Treatment policy	HbA1c (%)		ADLB		CHG		SEX BMIG3BL STRAT1V STRAT2V
ANCOVA - RD-MI	Treatment policy	HbA1c (mmol/mol)		ADLB		CHG		SEX BMIG3BL STRAT1V STRAT2V


Intermediate Program
Intermediate Program
Intermediate Program
Intermediate Dataset
Intermediate Dataset
Intermediate Dataset
Final Stacked Dataset
mk900lrtpmibw5
mk900mmrm1bw
mk900mmrm2bw
mk900rdmibiohscrp
mk900rdmibiopai1c
mk900rdmibw
mk900rdmidbp
mk900rdmifpg
mk900rdmifpgsi
tmmrm1bw
tmmrm2bio_hscrp
tmmrm2bio_lept
tmmrm2bio_pai1_c
tmmrm2bio_slept
tmmrm2bmi
tmmrm2bw
mk905axcomb
axcat.xpt
axcont.xpt
axstrat.xpt
axtp.xpt

Define.xml presentation

Data References (incl. Selection Criteria)	ADVS [FASFL = "Y" and ANL01FL = "Y" and ANELFL = "Y" and PARAM IN ("Body Fat Mass (kg)", "Body Fat Percentage (%)", "Lean Body Mass (kg)", "Lean Body Mass in % (%)", "Visceral Fat Mass (kg)", "Visceral Fat Mass in % (%)")] AXCONT [ESTIMAND = "TREATMENT POLICY" and POPVAR = "DXAFL" and MODEL = "ANCOVA - J2R-MI" and PARAM IN ("Body Fat Mass (kg)", "Body Fat Percentage (%)", "Lean Body Mass (kg)", "Lean Body Mass in % (%)", "Visceral Fat Mass (kg)", "Visceral Fat Mass in % (%)") and CHGTYPE = "CHG"] Results presented in table are selected from the AX data set. Results in the data set AX are based on imputation/modelling of data in the AD data set.

Programming Statements	[SAS version 9.4] Imputation of missing data and modelling/estimation was done using mk9xxx.txt referenced below. Estimates were added to the AXCONT data set, wherefrom estimates were selected when presenting estimates in the table using tstatancova.txt. Imputation of missing AVAL was done using PROC MI. Imputation was based on available values within imputation group where each combination of treatment arm and treatment status (on-/off-drug status) at landmark visit defined an imputation group. Last available observation on treatment is denoted LAO_OT. proc mi data = {data with missing results} out = {output data set} nimpute= 1000; by TRTP {treatment status}; class SEX BMIG3BL {LAO_OT time interval}; var SEX BMIG3BL {LAO_OT time interval} {AVAL at baseline} LAO_OT AVAL; monotone regression (AVAL = SEX BMIG3BL {LAO_OT time interval} {AVAL at baseline} LAO_OT); run; Modelling of change and consecutive estimation was done using PROC MIXED. proc mixed data = {data with missing replaced by imputed results}; by PARAM _IMPUTATION_; class TRTP; model CHG = TRTP {AVAL at baseline}; lsmeans TRTP/diff=control('Placebo') obsmargins at means cl; run; Results presented in table: proc print (using selection as specified in Data References); var MGROUP CGROUP NOBSPOP NOBST EST CFB CONTR CONTRLL CONTRUL PROBSUP; run; mk900j2rmidxabw.txt mk900j2rmidxabwkg.txt mk900j2rmidxalbm.txt mk900j2rmidxalbmkg.txt mk900j2rmidxatfm.txt mk900j2rmidxatfmkg.txt mk900j2rmidxavfm.txt mk900j2rmidxavfmkg.txt tstatancova.txt

Combined ADRG and define.xml templates

Review of define.xml and ADRG is part of developing and reviewing the ADaM program

One template project document to be used for all trials

Integrated summaries (ISS, ISE and pro questionnaires)

- Subgroup analysis made very easy due to a solid integrated database foundation
- Most of the programs are re-used from trial level, both the output and the statistical analysis

Opportunities and challenges

Doing analyses across the projects - challenges

Increased maintenance and rigidness – single updates required a lengthy process

Different definitions for Trial nr. 4 variables forced many changes in both ADaM and TFL

Quality of TFL programming could be improved

3 separate analysis pools, each with their own set of variables: phase3a, dose escalation, non-diabetic

Keeping everyone in the loop was critical

Adding new trial to the existing pools required general updates – loosing backwards compatibility

PMDA submission requirements different from FDA

Different requirements for the PMDA submission

FDA:

Table 1-2 Versions of standards and dictionaries used

Standard or Dictionary	Version(s) Used
SDTM	SDTM v1.4 SDTM IG v3.2.
ADaM	ADaM Model Document 2.1 ADaM Implementation Guide v1.1 ADaM Structure for Occurrence Data (OCCDS) 1.0
Controlled Terminology	ADaM 2018-12-21
Data Definitions	define.xml v2.0

PMDA:

Table 1-2 Versions of standards and dictionaries used

Standard or Dictionary	Version(s) Used
SDTM	SDTM v1.4 SDTM IG v3.2.
ADaM	ADaM Model Document 2.1 ADaM Implementation Guide v1.0 ADaM Structure for Occurrence Data (OCCDS) 1.0
Controlled Terminology	ADaM 2018-12-21
Data Definitions	define.xml v2.0

Things that were appreciated:

A large number of high quality, submission ready programs have been created

Repetitive programming has been largely eliminated

Making subgroups or slight modifications to programs and reviewing them was quite easy, consistent and transparent

Analysis Results Metadata was describing all programs used to create an individual output, making life easy for the reviewers

We have been asked multiple times during the Q&A to provide programs we've created in order to anwser the reviewer's questions. Having all our programs at a high 'submission-ready' level made this much easier

Outcome

FDA response to a pre-NDA meeting: ”Yes, your proposal appears acceptable. However, we may request additional datasets or programs if needed during the review process.”

Submission made without any technical issues

No questions to Biostatistics deliverables in the Q&A phase – neither in the FDA nor the PMDA side!

The Project was a team effort of many skilled Programmers and Statisticians – thank you all!

Thank You!

Questions?