

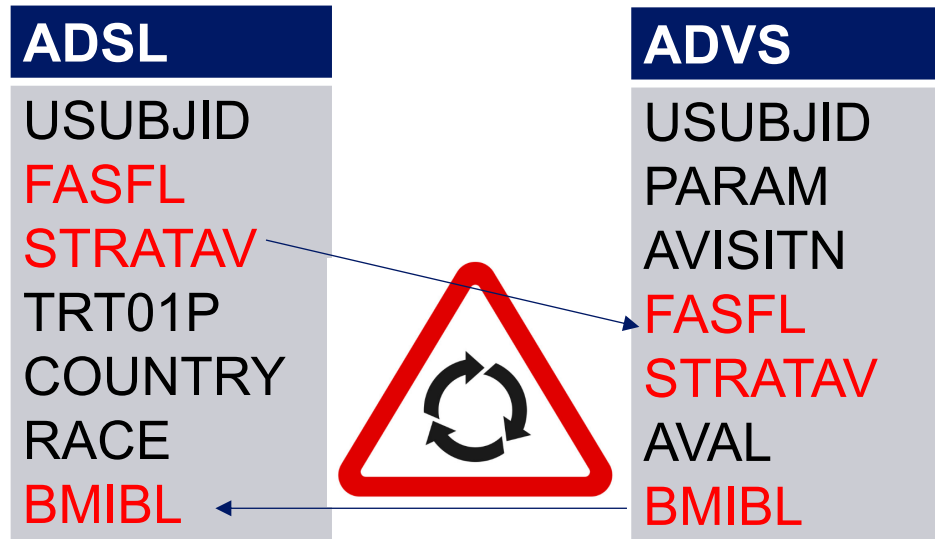
DH11: Flow Dependencies in ADaM Generation

Jeppe Juul, jxju@novonordisk.com

Principal Programmer

Motivation

- Some variables are present in (almost) all ADaM datasets



- The natural place to derive these variables vary:
 - Population flags such as FASFL: typically derived in ADSL but are also needed in e.g. ADVS
 - Baseline values such as BMIBL: typically derived in e.g. ADVS but are also needed in ADSL
 - Risk of circular flow!*

Key questions

- In which order should ADaM datasets be created?
- How can the selection of core variables be changed without modifying all ADaM programs?

ADSL

USUBJID
FASFL
~~STRATAV~~
STRATAR
TRT01P
COUNTRY
RACE
BMIBL

ADVS

USUBJID
PARAM
AVISITN
AVAL
FASFL
~~STRATAV~~
STRATAR
BMIBL

Why have the same variables in many datasets?

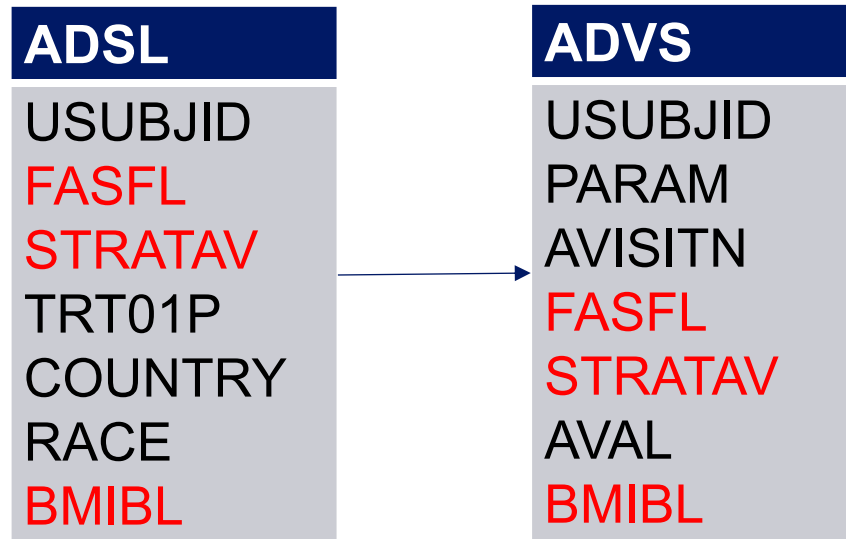
- Convenient to have the variables where they are needed
 - Fewer dataset merges ("one proc away")
- ADaM implementation guide 1.2:

"... at a minimum, any ADSL variable needed to enable analysis (e.g., statistical model covariates, population flags, subgrouping variables) should appear in the analysis dataset."
- FDA technical conformance guide:

"Core variables, which include covariates presented in the study protocol that are necessary to analyze data, should be included in each ADaM dataset"

In which order should
ADaM datasets be created?

Strategy 1: Derive all core variables in ADSL



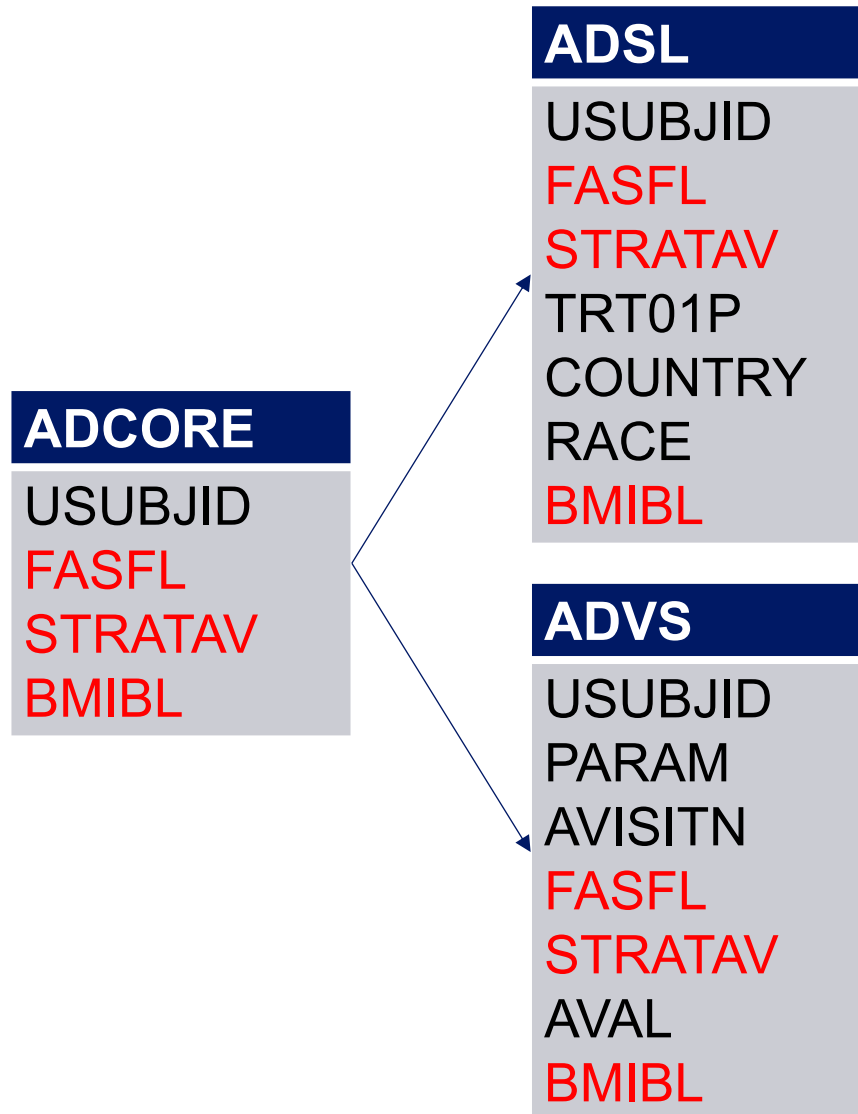
What tempted us:

- Simplest flow possible

So why not?:

- Hard to derive e.g. BMIBL without ADVS (may depend on visit reallocation, analysis flags, imputation rules, etc)

Strategy 2: First derive core variables in ADCORE



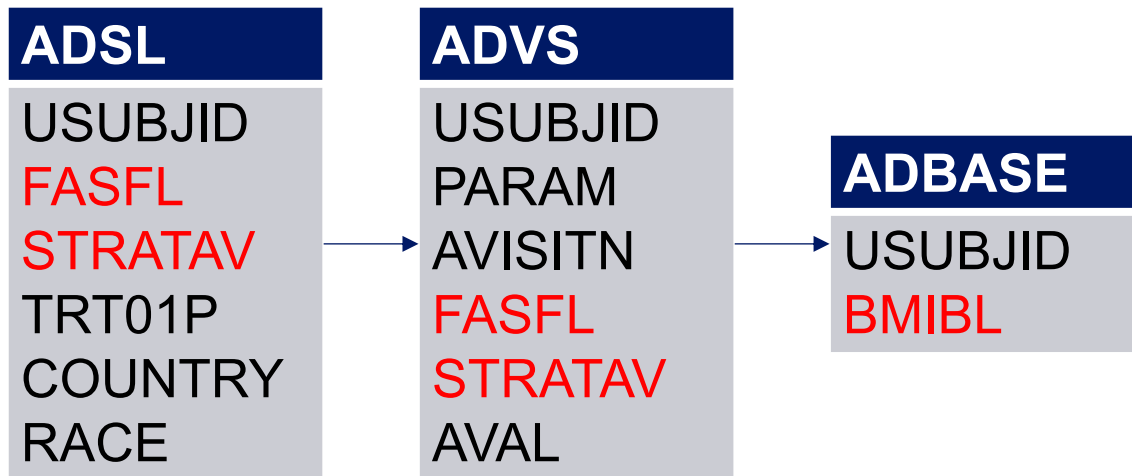
What tempted us:

- Simple data flow
- Minimal data dependency (e.g. no need to re-review ADVS after an update to ADSL)

So why not?:

- Hard to derive e.g. BMIBL without ADVS (may depend on visit reallocation, analysis flags, imputation rules, etc)

Strategy 3: Derive baseline variables last



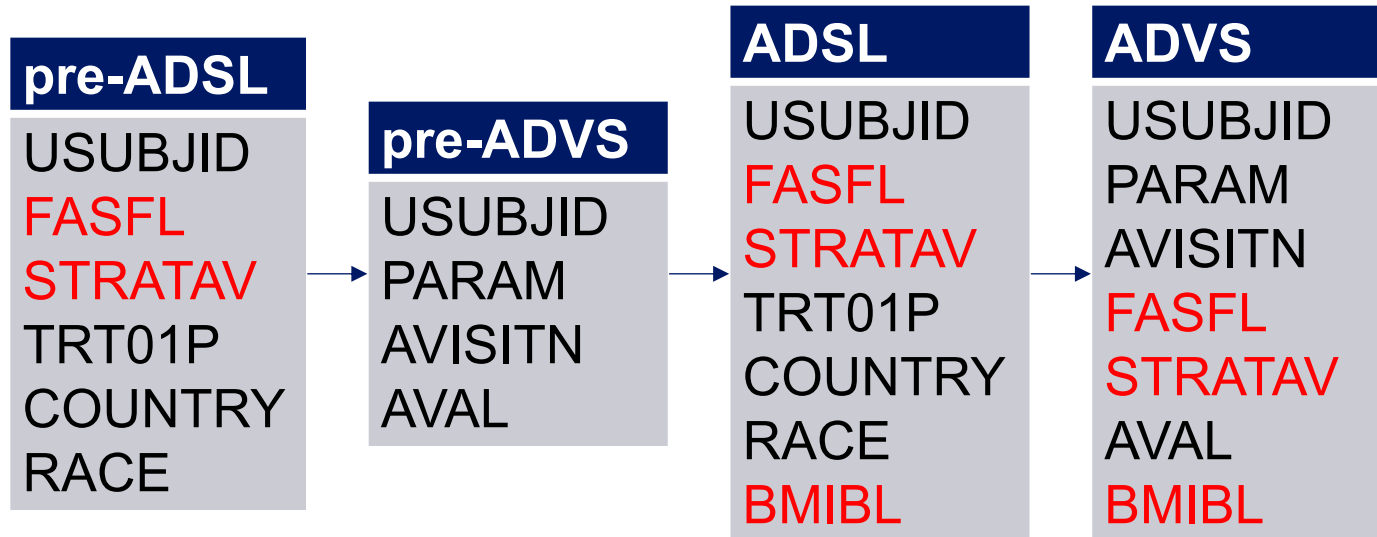
What tempted us:

- Variables are derived in the most convenient order
- Simple, linear flow

So why not?:

- Inconvenient to always merge with ADBASE
- Core variable not present in all datasets as required

Strategy 4: Two-step ADSL creation



Why we chose this strategy:

- Variables are derived in the most convenient order
- Core variables present in all datasets

What still bothers us:

- Intermediate "pre-" datasets overwritten during flow
- Always remember to run the last "add-core-variables" program!

Note: Strategy mentioned in ADaM Implementation Guide v1.2 sec 3.2

How can the selection of core variables be changed without modifying all ADaM programs?

Illustrating the problem

- Imagine STRATAV is added from ADSL in all other ADaM programs:

```
* Add core variables from ADSL;  
data advs_with_core;  
  merge advs_without_core      stratar  
        adsl(keep=usubjid strata);  
  by usubjid;  
run;
```

- If it is decided that STRATAR should be core instead then **all** ADaM programs must be updated:



Solution: Metadata driven approach

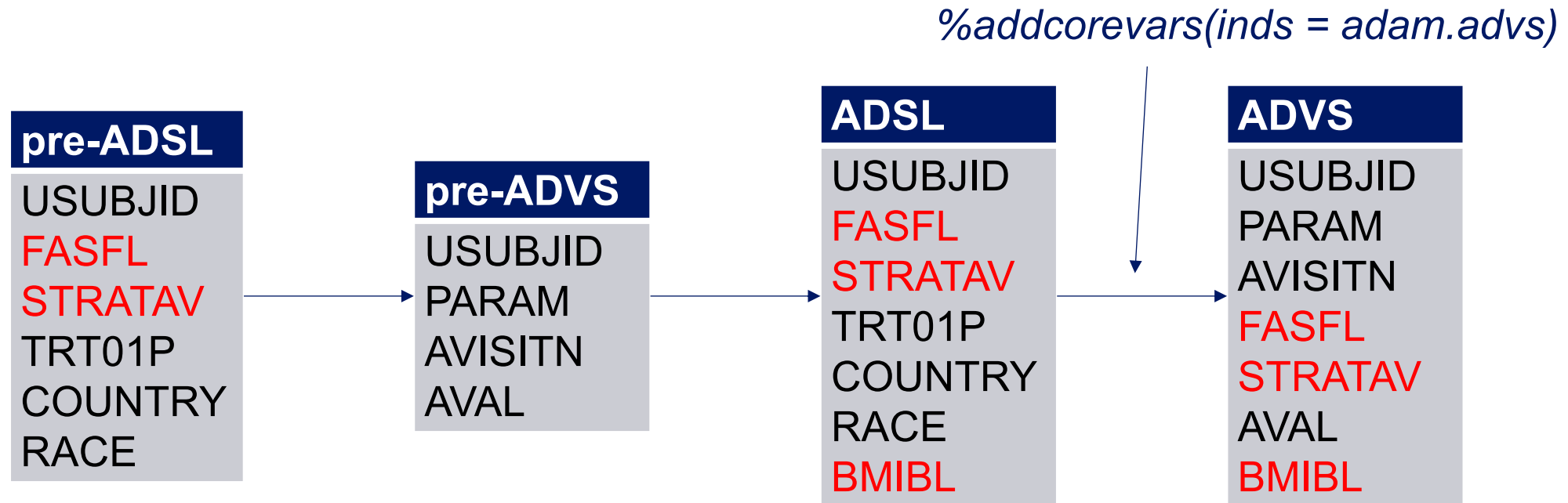
- We create a metadata dataset listing the core variables:

 TABLE	 COLUMN	 LABEL	 COREFL
ADSL	USUBJID	Unique Subject Identifier	
ADSL	FASFL	Full Analysis Set Population Flag	Y
ADSL	STRATAV	Strata from Verification Source	Y
ADSL	TRT01P	Planned Treatment for Period 01	
ADSL	COUNTRY	Country	
ADSL	RACE	Race	
ADSL	BMIBL	Body Mass Index at Baseline	Y

- Note: Similar dataset already exists in Novo Nordisk with metadata from define.xml used for adding labels, lengths, formats etc to ADaM dataset

Solution: Metadata driven approach

- An "add core variables" macro is applied to all adam datasets in the last step



Solution: Metadata driven approach

- The "add core variables" macro does the following when applied to e.g. ADVS:
 1. Extract list of core variables from metadata
 2. If any core variables already exist in ADVS these are dropped
 3. Read the core variables from ADSL
 4. Join ADVS and ADSL by USUBJID to add the core variables
- Consequence:
Selection of core variables can be changed without modifying **any** ADaM program!

Thank you for your attention!

Questions?: jxju@novonordisk.com

Acknowledgements

Niels-Kristian Kjøller

Anne Katrine Alsing

Ari Siggaard Knoph