



# **“It’s ARM, Jim, But Not As We Know It” – Our Journey to Automating Output Delivery Through Metadata**

Edwin van Stein

# Agenda

1. Who's Jim?
2. Why and Our Vision
3. Definition of Analysis
4. Analysis Concepts
5. Analysis Results Metadata
6. Analysis Results Data
7. Analysis Output Metadata
8. Next steps

## Disclaimer

This is very much a work in progress. By the time I'm actually presenting this to you, our internal version will look different.

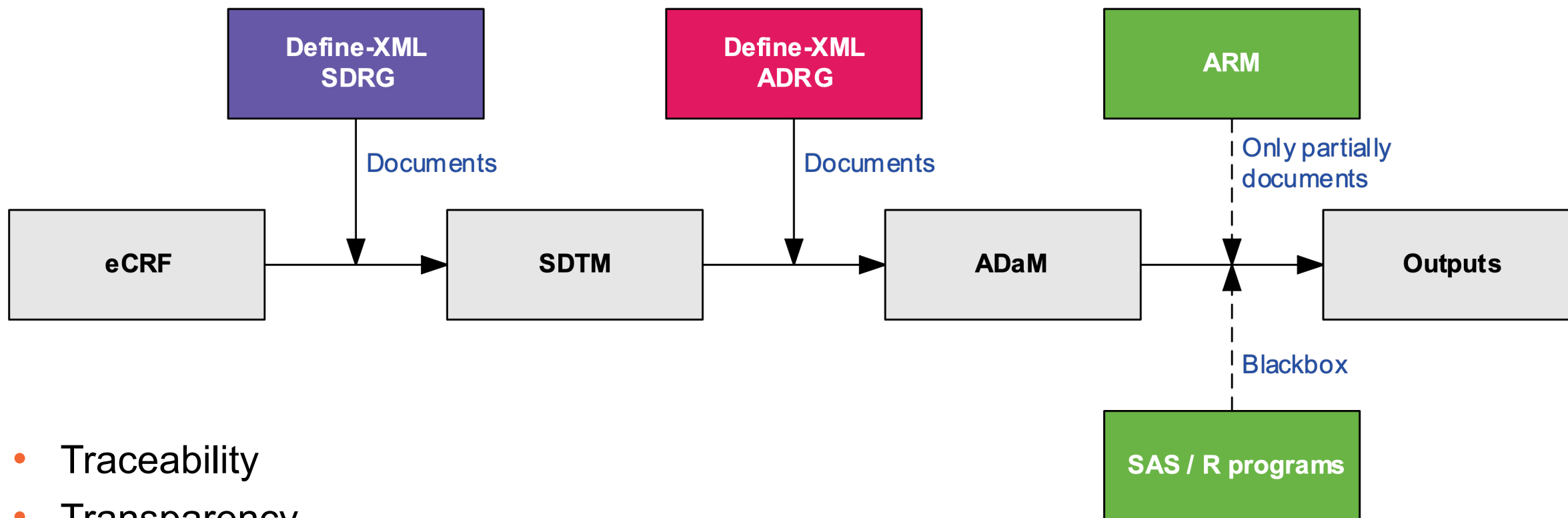
## Who's Jim?

“It's ARM, Jim, But Not As We Know It”

“It's **Life**, Jim, But Not As We Know It”



# Why and Our Vision



- Traceability
- Transparency
- Automation
- Consistency

# Why and Our Vision

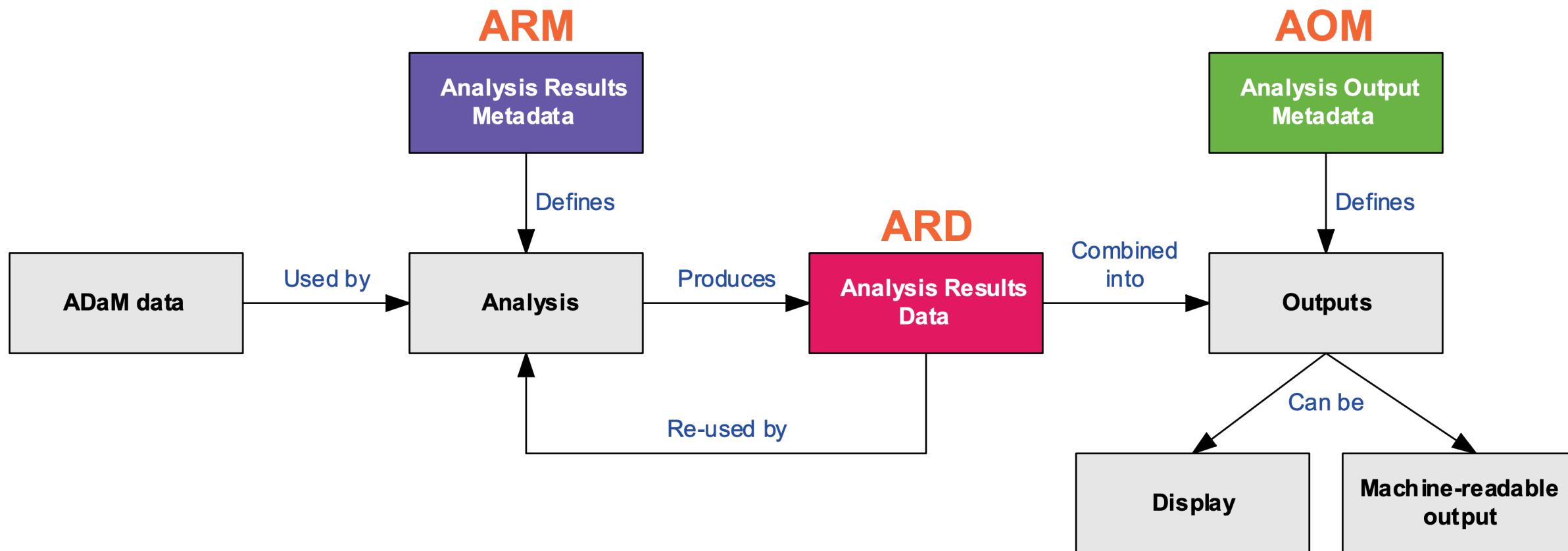
## WORM

**Write Once, Read Many**

- Any analysis defined once
- Any analysis executed once
- Any analysis validated once
- Re-use analyses across outputs
- Re-use analyses across analyses



# Why and Our Vision



# Definition of Analysis

A single analysis on one variable using one method

- Frequency counts by treatment (e.g., big N)
- Frequency counts by treatment and sex
- Frequency counts by treatment, sex and age group
- Summary statistics on age by treatment
- Summary statistics on age by treatment and sex

## Reusability

- How often do we report big N?
- How often do we use big N as denominator?
- Should we generate big N every time we need it?

| Table X: Summary of Demographic and Baseline Characteristics |                        |                        |                  |
|--------------------------------------------------------------|------------------------|------------------------|------------------|
|                                                              | Treatment A<br>(N=XXX) | Treatment B<br>(N=XXX) | Total<br>(N=XXX) |
| Sex                                                          |                        |                        |                  |
| n                                                            | xxx                    | xxx                    | xxx              |
| F                                                            | xx (xx%)               | xx (xx%)               | xx (xx%)         |
| M                                                            | xx (xx%)               | xx (xx%)               | xx (xx%)         |
| Age (YEARS) at First Dose                                    |                        |                        |                  |
| n                                                            | xxx                    | xxx                    | xxx              |
| Mean                                                         | xx                     | xx                     | xx               |
| SD                                                           | xx                     | xx                     | xx               |

# Analysis Concepts

- Information is needed multiple times:
  - By variables like TRTAN and TRTA
  - Safety Analysis Set is always SAFFL="Y"
- Information is linked:
  - Wherever TRTA is used, TRTAN should be used to order
- Analysis Variable Groups
- Analysis Variable Matrix
- Analysis Range Checks

# Analysis Concepts

## Analysis Variable Groups

- To link variables, their units and codelists
- Used across other Concepts, ARM and AOM

| Variable GroupID | Variable GroupLabel | Variable Num | Variable Code | Variable Decode | Variable Unit | Codelist |
|------------------|---------------------|--------------|---------------|-----------------|---------------|----------|
| TRTAN_TRTA       | Actual Treatment    | TRTAN        |               | TRTA            |               | TRT      |
| TRTPN_TRTP       | Planned Treatment   | TRTPN        |               | TRTP            |               | TRT      |
| ATOXGR_ATOXGRN   | Toxicity Grade      | ATOXGRN      |               | ATOXGR          |               | ATOXGR   |
| AAGE1_AAGE1U     | Age at First Dose   | AAGE1        |               |                 | AAGE1U        |          |
| SEXN_SEX         | Sex                 | SEXN         | SEX           |                 |               | SEX      |
| PARAMN_PARAM     | Parameter           | PARAMN       | PARAMCD       | PARAM           |               | PARAM    |
| AVISITN_AVISIT   | Visit               | AVISITN      |               | AVISIT          |               | AVISIT   |
| ATPTN_ATPT       | Timepoint           | ATPTN        |               | ATPT            |               | ATPT     |

# Analysis Concepts

## Analysis Variable Matrix

- To define (combinations of) variable(s) (groups)
- Used across ARM and AOM

| Variable MatrixID | VariableMatrixLabel            | Variables                                                        | FullMatrix |
|-------------------|--------------------------------|------------------------------------------------------------------|------------|
| AVISIT_ATPT       | Visit, Timepoint               | AVISITN_AVISIT , ATPTN_ATPT                                      | N          |
| SEX_FULL          | Sex (full matrix)              | [Visit](local:AVISITN_AVISIT) ,<br>[Timepoint](local:ATPTN_ATPT) |            |
| TRTA_FULL         | Actual Treatment (full matrix) |                                                                  |            |
| SOC               | System Organ Class             | AESOC                                                            | N          |
| PT                | Preferred Term                 | AEDECOD                                                          | N          |

# Analysis Concepts

## Analysis Range Checks

- To define range checks
- Used across ARM and AOM

| Range CheckID | RangeCheckLabel     | RangeCheck        |
|---------------|---------------------|-------------------|
| SAF           | Safety Analysis Set | SAFFL EQ "Y"      |
| TRTEM         | Treatment Emergent  | TRTEMFL EQ "Y"    |
| SEX_MF        | Sex (Male, Female)  | SEX IN ("M", "F") |

- CDISC compliant
  - Only EQ, GT, GE, LT, LE, NE, IN, NOTIN
  - Only AND operator
  - Only parenthesis for IN and NOTIN
- More complexity needed? Define additional variables/flags in ADaM

# Analysis Results Metadata

## Duplication ARM

- Examples:
  - Duplicating all records to produce a total column
  - Duplicating certain records to produce aggregate categories
- Don't always want to do duplication in ADaM

| Duplication GroupID | Variables      | InputNum | InputCode | DuplicatedNum | DuplicatedCode | DuplicatedDecode |
|---------------------|----------------|----------|-----------|---------------|----------------|------------------|
| TRTA_TOT            | TRTAN_TRTA     | 1        |           | 9             |                | Total            |
| TRTA_TOT            | TRTAN_TRTA     | 2        |           | 9             |                | Total            |
| ATOXGR_345          | ATOXGRN_ATOXGR | 3        |           | 8             |                | Grade 3+4+5      |
| ATOXGR_345          | ATOXGRN_ATOXGR | 4        |           | 8             |                | Grade 3+4+5      |
| ATOXGR_345          | ATOXGRN_ATOXGR | 5        |           | 8             |                | Grade 3+4+5      |

- Duplications are applied sequentially

# Analysis Results Metadata

## Main ARM

- One record defines one analysis
- Refers to Analysis Concepts and Duplication ARM

| Table X: Summary of Demographic and Baseline Characteristics |                        |                        |                  |
|--------------------------------------------------------------|------------------------|------------------------|------------------|
|                                                              | Treatment A<br>(N=XXX) | Treatment B<br>(N=XXX) | Total<br>(N=XXX) |
| Sex                                                          |                        |                        |                  |
| n                                                            | xxx                    | xxx                    | xxx              |
| F                                                            | xx (xx%)               | xx (xx%)               | xx (xx%)         |
| M                                                            | xx (xx%)               | xx (xx%)               | xx (xx%)         |
| Age (YEARS) at First Dose                                    |                        |                        |                  |
| n                                                            | xxx                    | xxx                    | xxx              |
| Mean                                                         | xx                     | xx                     | xx               |
| SD                                                           | xx                     | xx                     | xx               |

| Variable           | Value                                             |
|--------------------|---------------------------------------------------|
| ARMID              | AS_SAF                                            |
| DuplicationGroupID | TRTA_TOT                                          |
| Domain             | ADSL                                              |
| AnalysisSet        | [Safety Analysis Set](local:SAF)                  |
| Method             | CountDistinctSubject                              |
| ByVariables        | [Actual Treatment (full matrix)](local:TRTA_FULL) |

- Tool agnostic: methods, not (pseudo-)code

# Analysis Results Metadata

## Main ARM

| Table X: Summary of Demographic and Baseline Characteristics |                        |                        |                  |
|--------------------------------------------------------------|------------------------|------------------------|------------------|
|                                                              | Treatment A<br>(N=XXX) | Treatment B<br>(N=XXX) | Total<br>(N=XXX) |
| Sex                                                          |                        |                        |                  |
| n                                                            | xxx                    | xxx                    | xxx              |
| F                                                            | xx (xx%)               | xx (xx%)               | xx (xx%)         |
| M                                                            | xx (xx%)               | xx (xx%)               | xx (xx%)         |
| Age (YEARS) at First Dose                                    |                        |                        |                  |
| n                                                            | xxx                    | xxx                    | xxx              |
| Mean                                                         | xx                     | xx                     | xx               |
| SD                                                           | xx                     | xx                     | xx               |

| Variable           | Value                              |
|--------------------|------------------------------------|
| ARMID              | DM_SEX_ANY                         |
| DuplicationGroupID | TRTA_TOT                           |
| Domain             | ADSL                               |
| AnalysisSet        | SAF                                |
| SubSet             | [Sex (Male, Female)](local:SEX_MF) |
| Method             | CountDistinctSubject               |
| ByVariables        | TRTA_FULL                          |

# Analysis Results Metadata

## Main ARM

Table X: Summary of Demographic and Baseline Characteristics

|                           | Treatment A<br>(N=XXX) | Treatment B<br>(N=XXX) | Total<br>(N=XXX) |
|---------------------------|------------------------|------------------------|------------------|
| Sex                       |                        |                        |                  |
| n                         | xxx                    | xxx                    | xxx              |
| F                         | xx (xx%)               | xx (xx%)               | xx (xx%)         |
| M                         | xx (xx%)               | xx (xx%)               | xx (xx%)         |
| Age (YEARS) at First Dose |                        |                        |                  |
| n                         | xxx                    | xxx                    | xxx              |
| Mean                      | xx                     | xx                     | xx               |
| SD                        | xx                     | xx                     | xx               |

### Variable

### Value

ARMID

DM\_SEX

DenomID

DM\_SEX\_ANY

DuplicationGroupID

TRTA\_TOT

Domain

ADSL

AnalysisSet

SAF

SubSet

SEX\_MF

Method

CountDistinctSubject

ByVariables

[Actual Treatment (full  
matrix)](local:TRTA\_FULL)  
\* [Sex (full matrix)](local:SEX\_FULL)

# Analysis Results Metadata

## Main ARM

| Table X: Summary of Demographic and Baseline Characteristics |                        |                        |                  |
|--------------------------------------------------------------|------------------------|------------------------|------------------|
|                                                              | Treatment A<br>(N=XXX) | Treatment B<br>(N=XXX) | Total<br>(N=XXX) |
| Sex                                                          |                        |                        |                  |
| n                                                            | xxx                    | xxx                    | xxx              |
| F                                                            | xx (xx%)               | xx (xx%)               | xx (xx%)         |
| M                                                            | xx (xx%)               | xx (xx%)               | xx (xx%)         |
| Age (YEARS) at First Dose                                    |                        |                        |                  |
| n                                                            | xxx                    | xxx                    | xxx              |
| Mean                                                         | xx                     | xx                     | xx               |
| SD                                                           | xx                     | xx                     | xx               |

| Variable           | Value                                   |
|--------------------|-----------------------------------------|
| ARMID              | DM_AGE                                  |
| DuplicationGroupID | TRTA_TOT                                |
| Domain             | ADSL                                    |
| AnalysisSet        | SAF                                     |
| Method             | SummaryStat                             |
| Analysis           | [Age at First Dose](local:AAGE1_AAGE1U) |
| ByVariables        | TRTA_FULL                               |

## ► Analysis Results Data

- Normalized structure for results
- Machine-readable (not necessarily human-readable)
- Not meant to be consumed without ARM

| ARMID  | QualifierID | Role       | Variable   | ValueNum | ValueCode | ValueDecode |
|--------|-------------|------------|------------|----------|-----------|-------------|
| AS_SAF | 1.01        | ByVariable | TRTAN:TRTA | 1        |           | Treatment A |
| AS_SAF | 1.01        | Statistic  |            | 49       | N         |             |
| AS_SAF | 2.01        | ByVariable | TRTAN:TRTA | 2        |           | Treatment B |
| AS_SAF | 2.01        | Statistic  |            | 51       | N         |             |
| AS_SAF | 3.01        | ByVariable | TRTAN:TRTA | 9        |           | Total       |
| AS_SAF | 3.01        | Statistic  |            | 100      | N         |             |

## Analysis Results Data

| ARMID  | QualifierID | Role       | Variable   | ValueNum      | ValueCode | ValueDecode |
|--------|-------------|------------|------------|---------------|-----------|-------------|
| DM_SEX | 1.01        | ByVariable | TRTAN:TRTA | 1             |           | Treatment A |
| DM_SEX | 1.01        | ByVariable | SEXN:SEX   | 1             | M         |             |
| DM_SEX | 1.01        | Statistic  |            | 25            | N         |             |
| DM_SEX | 1.02        | ByVariable | TRTAN:TRTA | 1             |           | Treatment A |
| DM_SEX | 1.02        | ByVariable | SEXN:SEX   | 1             | M         |             |
| DM_SEX | 1.02        | Statistic  |            | 49            | Denom     |             |
| DM_SEX | 1.03        | ByVariable | TRTAN:TRTA | 1             |           | Treatment A |
| DM_SEX | 1.03        | ByVariable | SEXN:SEX   | 1             | M         |             |
| DM_SEX | 1.03        | Statistic  |            | 51.0204081633 | Percent   |             |

## Analysis Results Data

| ARMID  | QualifierID | Role       | Variable   | ValueNum      | ValueCode | ValueDecode |
|--------|-------------|------------|------------|---------------|-----------|-------------|
| DM_SEX | 1           | ByVariable | TRTAN:TRTA | 1             |           | Treatment A |
| DM_SEX | 1           | ByVariable | SEXN:SEX   | 1             | M         |             |
| DM_SEX | 1.01        | Statistic  |            | 25            | N         |             |
| DM_SEX | 1.02        | Statistic  |            | 49            | Denom     |             |
| DM_SEX | 1.03        | Statistic  |            | 51.0204081633 | Percent   |             |

| QualifierID | TRTAN | TRTA        | SEXN | SEX | N  | Denom | Percent       |
|-------------|-------|-------------|------|-----|----|-------|---------------|
| 1           | 1     | Treatment A | 1    | M   | 25 | 49    | 51.0204081633 |
| 2           | 1     | Treatment A | 2    | F   | 24 | 49    | 48.9795918367 |
| 3           | 2     | Treatment B | 1    | M   | 25 | 51    | 49.0196078431 |
| 4           | 2     | Treatment B | 2    | F   | 26 | 51    | 50.9803921568 |
| 5           | 9     | Total       | 1    | M   | 50 | 100   | 50            |
| 6           | 9     | Total       | 2    | F   | 50 | 100   | 50            |

# Analysis Output Metadata

- Work in progress
- Will build upon metadata previously discussed, including concepts
- Defines an output
  - Display (table, figure, listing)
  - Machine-readable file (JSON, XML, etc. for disclosure, CSR automation, etc.)
- Defines:
  - Which sets of ARD need to be included (if any)
  - How they should be combined
  - And how to report them
- Tool agnostic: methods, not (psuedo-)code

# Analysis Output Metadata

- For a table AOM is split in two:
  - Transformation AOM:
    - Sets of ARD to combine
    - Subsetting
    - Sorting
    - Value formatting and alignment
    - Rows and groups
  - Reporting AOM
    - Titles and footnotes
    - By group processing (including subtitles)
    - Columns, labels and sorting
    - (Spanning) headers
    - Big N printing

## Next steps

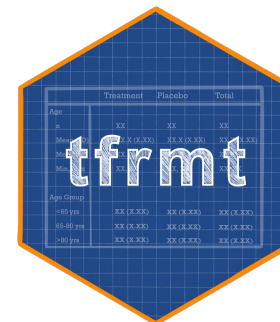
- Finalize ARM and ARD
- When producing ARD, store precision of numeric input variables
- Non-trivial statistics
- Further define AOM for tables
- Define AOM for listings
- Define AOM for figures
- Define AOM for machine-readable outputs

ClinicalTrials.gov



## Next steps

- Leveraging the metadata more
  - AE table that uses the **Actual Treatment** and **System Organ Class, Preferred Term** concepts as by variables, the **Treatment Emergent** and **Serious Adverse Event** range checks and the **Safety Analysis Set** concept for the analysis set... Do we really want to define the title manually?
- Cultural change / change management
- Tool development to ingest ARM to produce ARD is ongoing
- Tool development to ingest ARD and AOM to produce tables has started
  - Based on tfrmt (<https://gsk-biostatistics.github.io/tfrmt/>)
- MDR / GUI



## Next steps

- Using Biomedical Concepts...
- Linking it all together in CLD
  - Protocol
  - eCRF
  - SDTM
  - ADaM
  - ARM
  - ARD
  - AOM
  - Outputs

# Questions

Your comments and questions are valued and encouraged. Contact the author at:

Edwin van Stein

[edwin.j.stein@gsk.com](mailto:edwin.j.stein@gsk.com)

<https://linkedin.com/in/ejvanstein>



GSK