Development of ADaM creation tool
Towards future Automation

Ryo Nakaya, Fumihiro Yamasaki, Yohei Takanami
Takeda Pharmaceutical Company
@PharmaSUG SDE-Japan

Better Health, Brighter Future
Automation in the creation of statistical deliverables such as SDTM, ADaM, TFLs is the key to the development of the streamlined business process in drug development to reduce cost, save time, and maximize the quality and productivity. Recently many attempts can be seen in the industry.

- Automated generation tool
- low-code and hyper-automation solutions

We introduce one of our attempts to develop a tool that generates ADaM datasets with a standardized and streamlined process towards future automation as part of our business process internalization efforts, which has resulted in a reduced workload, time and cost while keeping high quality.
Motivation of Developing ADaM Creation Tool

Challenges in Outsourced Model

Specs
- Different spec formats by several vendors
- Unclear Define.xml creation process (no ad-hoc reproducitivity)
- Takeda Guidance of ADaM was created with general guidance (more review needed in studies) without machine readability for automation

Programs
- Manually writing programs
  - Inefficient
  - Error-prone
- Black box, that is, Difficulty in modifying and re-running programs without macros of vendors’ intellectual property during e-Data preparation and submission process

Transition from Outsource model to In-house model

Needs to have templates/tools in order to **create statistical deliverables in-house** as well as **to better manage cases of outsourced models.**
Agenda

• Statistical Deliverables and Automation in Clinical Development
• Semi-automated ADaM Creation Tool in Takeda
• Use cases, Lessons Learned and Challenges
• Summary and Next Step
Agenda

• Statistical Deliverables and Automation in Clinical Development
• Semi-automated ADaM Creation Tool in Takeda
• Use cases, Lessons Learned and Challenges
• Summary and Next Step
Statistical Deliverables and Automation in Clinical Development

Analysis document
created based on protocol. Summaries of major analysis plan in protocol and further details need to be added in SAP. A consideration on how much complete or machine readable for ADaM/TFL creation is necessary.

Creation of analysis datasets
using SAP and source SDTM datasets. There are common datasets across studies and therapeutic areas while additional datasets specific to study/therapeutic area.

Mapping
from various source data including EDC, lab, other external data is needed. Less derivation than ADaM creation.

Format
of analysis outputs is important. Using Standardized TFL shell can bring about highly re-usable programs/macros and reduce programming workload.
Agenda

- Statistical Deliverables and Automation in Clinical Development
- Semi-automated ADaM Creation Tool in Takeda
- Use cases, Lessons Learned and Challenges
- Summary and Next Step
Semi-automated ADaM Creation Tool in Takeda

- **SAS Macros**
- **ADaM Specs**
- **ADaM Standard Library**

- Semi-automated code generation
- ADaM Spec Templates for Takeda in Excel format used for each study
- Guidance for Takeda ADaM Standard
The tool

- Consists of ADaM Spec Templates and SAS Macros
- Has Excel formatted templates with default values for Common(Base) part of ADaM
- Has macros that automatically generate SAS codes by reading the spec
- Generates define.xml using *tsClinical Define.xml Generator©

*product of Fujitsu
As of Oct2023, the tool is available as part of tsClinical Metadata Desktop Tools. https://github.com/tsClinical/tsc-desktop
Process flow of Creating Spec and Programs

1. Start
2. Create aCRF
3. Create SDTM Specs
4. Create SAP
5. Create ADaM Specs
6. Create ADaM Specs
7. Create Wrapper Programs
8. End

Less effort by using spec template

Less manual programming by calling Macros with data-driven approach
Process flow of Executing Programs to create Datasets

1. **Start**
2. **Execute Wrapper Programs**
   - ADaM Specs
   - SDTM datasets

3. **Call Macro Programs**
   - **ADaM datasets**
   - **ADSL**

4. **Execute**
   - **ADaM datasets**

5. **End**

**Study/therapeutic specific additional coding.**

---

```sas
* ============== Setting ;
options sasautos="C:\%work\%データ加工共有\%データ加工 misc logic symbolsen nofmterr ;
libname INDATA "C:\%Users\%005138\%OneDrive - Takeda" ;
libname OUTDATA "C:\%Users\%005138\%OneDrive - Takeda" ;

%m_adae : 
* ============== Standard Macro ;

*============= Non-standard Processing ;
data WK_ADAEMNS_OUT ;
set MNK_ADAEM_OUT ;
select ;
   when ( AEAFAODM="N" and AETRTEM="N" )
   when ( AEAFAODM="Y" and AETRTEM="N" )
   when ( AEAFAODM="Y" and AETRTEM="Y" and aastdtl )
else ;
end ;
run ;
```
Spec Templates

Cover common datasets

- ADSL
- ADAE, ADCM, ADMH, ADDV (OCCDS)
- ADLB, ADVS, ADEG, ADPC, ADPP (BDS)

tsClinical Define.xml Generator© based spec. Moved and added some columns and sheets for efficiency in creation and review.

Users to modify sample logic (only if needed) and add study specific variables.

Cover common variables / common derivation logic (sample logic)

<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>Variable Name</th>
<th>Label</th>
<th>Key Seq</th>
<th>Vari</th>
<th>OdataType</th>
<th>CodeList</th>
<th>DisplayForm</th>
<th>Origin</th>
<th>Has Value</th>
<th>Missing Data</th>
<th>Predecessor/Derivation</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADSL</td>
<td>STUDYID</td>
<td>Study Identifier</td>
<td>text</td>
<td>Predecessor</td>
<td>Predecessor</td>
<td>= DM::STUDYID</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>SUBJID</td>
<td>Unique Subject Identifier</td>
<td>text</td>
<td>Predecessor</td>
<td>Predecessor</td>
<td>= DM::SUBJID</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>SITEID</td>
<td>Study Site Identifier</td>
<td>text</td>
<td>Predecessor</td>
<td>Predecessor</td>
<td>= DM::SITEID</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>COUNTRY</td>
<td>Country</td>
<td>text</td>
<td>Predecessor</td>
<td>Predecessor</td>
<td>= DM::COUNTRY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>AGU</td>
<td>Age Units</td>
<td>text</td>
<td>Predecessor</td>
<td>Predecessor</td>
<td>= DM::AGEU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>AGEU</td>
<td>Age</td>
<td>integer</td>
<td>Assigned</td>
<td>Numeric code for COUNTRY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>AGEDGR1</td>
<td>Pooled Age Group 1</td>
<td>text</td>
<td>Derived</td>
<td>Numeric code for AGEDGR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>AGEDGR1N</td>
<td>Pooled Age Group 1 (N)</td>
<td>integer</td>
<td>Derived</td>
<td>Numeric code for AGEDGR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>AGEDGR2</td>
<td>Pooled Age Group 2</td>
<td>text</td>
<td>Derived</td>
<td>Numeric code for AGEDGR2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>AGEDGR2N</td>
<td>Pooled Age Group 2 (N)</td>
<td>integer</td>
<td>Derived</td>
<td>Numeric code for AGEDGR2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>AGEDGR3</td>
<td>Pooled Age Group 3</td>
<td>text</td>
<td>Derived</td>
<td>[Takeda Comment]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>AGEDGR3N</td>
<td>Pooled Age Group 3 (N)</td>
<td>integer</td>
<td>Derived</td>
<td>[Takeda Comment]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>SEX</td>
<td>Sex</td>
<td>text</td>
<td>Predecessor</td>
<td>Predecessor</td>
<td>= DM::SEX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>SEXN</td>
<td>Sex (N)</td>
<td>integer</td>
<td>Assigned</td>
<td>Numeric code for SEX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>RACE</td>
<td>Race</td>
<td>text</td>
<td>Predecessor</td>
<td>Predecessor</td>
<td>= DM::RACE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL</td>
<td>RACEN</td>
<td>Race (N)</td>
<td>integer</td>
<td>Assigned</td>
<td>Numeric code for RACE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SAS Macros

Cover common datasets

- ADSL
- ADAE, ADCM, ADMH, ADDV (OCCDS)
- ADLB, ADVS, ADEG (BDS)

Features and functionality of Macros

- Dataset-level core macros consisting of small functional macros with better maintainability
- Framework of ADaM datasets is to be built by the core macros (incl. parameter, time window)
- Metadata-driven code creation is used for common derivations

[Diagram illustrating the relationships between ADSL, ADAE, and ADLB datasets with specific features such as parameter (PARAM) and time window (AVISIT).]
### Metadata-driven Code Creation for Common Derivations

Describe derivation in pre-defined rules to utilize metadata-driven functionality

Core macros read spec file and generate SAS codes

<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>Variable Name</th>
<th>Label</th>
<th>Key Sequence Code</th>
<th>DataTy</th>
<th>CodeList</th>
<th>DisplayForm</th>
<th>Origin</th>
<th>Has Value</th>
<th>Macro Processed</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADSL STUDIID</td>
<td>Study Identifier</td>
<td>text</td>
<td>text</td>
<td>Predecessor</td>
<td>1 DM.STUDID</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL USUBIID</td>
<td>Unique Subject Identifier</td>
<td>text</td>
<td>text</td>
<td>Predecessor</td>
<td>1 DM.USUBID</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL SUBIID</td>
<td>Subject Identifier for the Study</td>
<td>text</td>
<td>text</td>
<td>Predecessor</td>
<td>1 DM.SUBIID</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL SITEID</td>
<td>Study Site Identifier</td>
<td>text</td>
<td>text</td>
<td>Predecessor</td>
<td>1 DM.SITEID</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL COUNTRY</td>
<td>Country</td>
<td>text</td>
<td>COUNTRY</td>
<td>Predecessor</td>
<td>1 DM.COUNTRY</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL AGE</td>
<td>Age</td>
<td>integer</td>
<td>integer</td>
<td>Predecessor</td>
<td>1 DM.AGE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL AGEU</td>
<td>Age Units</td>
<td>text</td>
<td>AGEU</td>
<td>Predecessor</td>
<td>1 DM.AGEU</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL AGEGR1</td>
<td>Pooled Age Group 1</td>
<td>text</td>
<td>AGEGR1</td>
<td>Derived</td>
<td>1 DM.AGEGR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL AGEGR1N</td>
<td>Pooled Age Group 1 (N)</td>
<td>integer</td>
<td>AGEGR1N</td>
<td>Assigned</td>
<td>1 DM.AGEGR1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL AGEGR2</td>
<td>Pooled Age Group 2</td>
<td>text</td>
<td>AGEGR2</td>
<td>Derived</td>
<td>1 DM.AGEGR2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL AGEGR2N</td>
<td>Pooled Age Group 2 (N)</td>
<td>integer</td>
<td>AGEGR2N</td>
<td>Assigned</td>
<td>1 DM.AGEGR2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL AGEGR3</td>
<td>Pooled Age Group 3</td>
<td>text</td>
<td>AGEGR3</td>
<td>Derived</td>
<td>1 DM.AGEGR3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL AGEGR3N</td>
<td>Pooled Age Group 3 (N)</td>
<td>integer</td>
<td>AGEGR3N</td>
<td>Assigned</td>
<td>1 DM.AGEGR3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL SEX</td>
<td>Sex</td>
<td>text</td>
<td>SEX</td>
<td>Predecessor</td>
<td>1 DM.SEX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL SEXN</td>
<td>Sex (N)</td>
<td>integer</td>
<td>SEXN</td>
<td>Assigned</td>
<td>1 DM.SEX</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL RACE</td>
<td>Race</td>
<td>text</td>
<td>RACE</td>
<td>Predecessor</td>
<td>1 DM.RACE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADSL RACEN</td>
<td>Race (N)</td>
<td>integer</td>
<td>RACEN</td>
<td>Assigned</td>
<td>1 DM.RACE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A column to show variables to be processed by macro
### Example of derivations

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Description</th>
<th>Examples</th>
<th>Generated Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset-name.Variable-name ;</td>
<td>Assign values of the specified dataset and variable</td>
<td>= DM.AGE ;</td>
<td>data ADSL; merge ADSL DM(keep=USUBJID AGE rename=(AGE = _AGE)); by USUBJID; AGE = _AGE; run;</td>
</tr>
<tr>
<td>Dataset-name.Variable-name where Where-expression ;</td>
<td>Assign values of the specified dataset and variable after narrowing down to one-record-per-subject using extraction condition in where statement</td>
<td>= VS.VSSTRESN where VS.VSTESTCD = &quot;HEIGHT&quot; ;</td>
<td>data ADSL; merge ADSL VS(keep=USUBJID VSTESTCD VSSTRESN rename=(VSTESTCD = _VSTESTCD VSSTRESN = _VSSTRESN) where=(_VSTESTCD = &quot;HEIGHT&quot;)) ; by USUBJID ; HTBL = _VSSTRESN ; run ;</td>
</tr>
<tr>
<td>If Condition1 then Result1 ; else if Condition2 then Result2 ; else Result3 ;</td>
<td>Assign values using if/then statement. Select/when can be applied as well.</td>
<td>if . &lt; WTBL &lt; 50 then WTBLGR1 = &quot;Min&lt;= - &lt;50&quot; ; else if 50 &lt;= WTBL then WTBLGR1 = &quot;50&lt;= - &lt;=Max&quot; ;</td>
<td>data ADSL ; set ADSL ; if . &lt;= WTBL &lt; 50 then WTBLGR1 = ‘Min&lt;= - &lt;50’ ; else if 50 &lt;= WTBL then WTBLGR1 = '50&lt;= - &lt;=Max' ; run ;</td>
</tr>
<tr>
<td>Numeric code for Variable-name</td>
<td>Assign numeric code using codelist information for character variable</td>
<td>Numeric code for WTBLGR1</td>
<td>data ADSL ; set ADSL ; WTBLGR1N = input(WTBLGR1, WTBLGR1N.) ; run ; * WTBLGR1N format was created using CODELIST sheet in advance.</td>
</tr>
</tbody>
</table>
Agenda

- Statistical Deliverables and Automation in Clinical Development
- Semi-automated ADaM Creation Tool in Takeda
- Use cases, Lessons Learned and Challenges
- Summary and Next Step
Use cases, Lessons Learned and Challenges

Experiences in use cases

- Spec templates cover common datasets/variables
- Macros cover coding of 50-80% in ADSL, 70-90% in BDS variables, 90% or more in OCCDS

<table>
<thead>
<tr>
<th></th>
<th>ADSL</th>
<th>ADAE</th>
<th>ADLB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study AAA</td>
<td>45%</td>
<td>90%</td>
<td>90%</td>
</tr>
<tr>
<td>Study BBB</td>
<td>75%</td>
<td>92%</td>
<td>67%</td>
</tr>
<tr>
<td>Study CCC</td>
<td>82%</td>
<td>92%</td>
<td>80%</td>
</tr>
<tr>
<td>Study DDD</td>
<td>77%</td>
<td>100%</td>
<td>79%</td>
</tr>
<tr>
<td>Study EEE</td>
<td>87%</td>
<td>100%</td>
<td>96%</td>
</tr>
</tbody>
</table>

Learned cases tend to have smaller coverage
- External file for derivation
- Cyclic time window (especially in oncology)

Challenges

- Pinnacle format of spec is commonly used in global region
- CSV (Computerized System Validation) to be discussed
- Room to utilize generative AI tools in drafting study spec and code generation in macros (rule-based derivation can be more flexible)

*Cover rate is based on how many of variables did not require manual coding.
Agenda

- Statistical Deliverables and Automation in Clinical Development
- Semi-automated ADaM Creation Tool in Takeda
- Use cases, Lessons Learned and Challenges
- Summary and Next Step
With the industry standard, innovative technologies such as CDISC and AI/ML, the business process in the creation of statistical deliverables can be simplified and streamlined towards a fully automated process or system.

More than 80% of ADaM datasets/variables can be automatically generated using the ADaM creation tools and enhancement and improvement are underway.

Technology-based innovation needs a variety of knowledge to realize and keywords to achieve the innovation would be “Communication” and “Collaboration”.