Standardising The Standards: The Benefits of Consistency

Nathan James, Roche Products Ltd.
SDTM - An End User’s Perspective

- Improved submission of clinical data to Regulators
- Improved transmission of clinical data between CRO and Sponsor

Is the implementation of the SDTM model *always* the same?

**CLAIM:**

The interpretation and implementation of the model can vary between companies
A Simple Analogy

“Bring me a standard red apple!”
• SDTM build was *outsourced* for the project

• 2 studies had the SDTM construction with *2 different vendors*

- Future *pooling* of studies likely

- *Assumed* standard framework and structure, regardless of who was constructing the SDTM.
The Reality

- SDTM s differed in a variety of ways
- 2 categories:
  - Expected differences
  - Unexpected differences
Expected Differences

- There will always be expected differences
- Based on the similarity of the collected data
- Safety domains more comparable than efficacy domains
Expected Differences

1. Domains delivered
2. Study specific elements
3. Trial Design Domains
4. SDTM Versions
Expected Differences

STUDY 1

- DS
- DM
- AE
- EX
- LB

STUDY 2

- DS
- DM
- AE
- EX
- LB

- TA
- TE
- TI
- TS
- TV

- QS
- ZA

- TA
- TE
- TI
- TS
- TV

- RS
- TR
Expected Differences

1. Domains delivered

2. Study specific elements

3. Trial Design Domains

4. SDTM Versions
Expected Differences

1. Domains delivered
2. Study specific elements
3. **Trial Design Domains**
4. SDTM Versions
Expected Differences

1. Domains delivered
2. Study specific elements
3. Trial Design Domains
4. SDTM Versions
Unexpected Differences

1. Subjective interpretations of the model
2. Mapping decisions
Unexpected Differences

• Related to derivations performed within the model
  – Trial Design Domains
  – Exposure domain

• Input from an analysis perspective
  – Reference start date [DM.RFSTDTC]
  – Baseline flags [--BLFL]
Unexpected Differences

1. Subjective interpretations of the model

2. Mapping decisions
Unexpected Differences

• Data that has **no obvious mapping** into the model
  – *Deaths collected as an event*
In the case of patient’s death

Death, whether or not related to the study medication(s), must be immediately reported to sponsor by telefax on an SAE form.

Date of death

Primary cause of death

Underlying cause of death (if applicable)

Relationship of death to study medication

unrelated □
remote □
possible □
probable □

Was autopsy performed? yes □ no □

If yes, please summarize findings:

General comment:
Unexpected Differences

• Data that has no obvious mapping into the model
  – Deaths collected as an event
  – Non-standard elements

• No right or wrong way of mapping some data
  – Different vendors take different approaches
Automated Checking Tool

- **Extent** of the differences was unknown
- **Automated tool** to check for discrepancies developed
- **More benefits** to the tool than first perceived
Automated Checking Tool

REQUIREMENTS

✓ List **all discrepancies** between 2 sets of SDTM data

✓ Focus on **consistency** of tabulation and not conformance

✓ Dynamic

✓ Simple to use

```bash
.sdmt_consistence_check( study1  = /opt/BIOSTAT/sdtm/cdpt9999/xx11111
, study2  = /opt/BIOSTAT/sdtm/cdpt9999/xx22222
 );
```
Automated Checking Tool

DOMAIN CHECK

- Checks differences in the existence of domains

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>Member Name</td>
</tr>
<tr>
<td>2</td>
<td>DA</td>
</tr>
<tr>
<td>3</td>
<td>SE</td>
</tr>
<tr>
<td>4</td>
<td>XP</td>
</tr>
</tbody>
</table>

STUDY 1 (xx11111)

- DA

STUDY 2 (xx22222)

- SE
  - XP
Automated Checking Tool

VARIABLE CHECK

- Checks differences in the variables – existence and attributes
- Only common domains compared
## Automated Checking Tool

**Table 1**

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Member Name</td>
<td>Column Name</td>
<td>COMMENT</td>
<td>Column Type</td>
</tr>
<tr>
<td>1</td>
<td>AE</td>
<td>AEACN1</td>
<td>Domain AE, variable AEACN1 is in cdpt9999/xx22222 but not in cdpt9999/xx11111</td>
<td>char</td>
</tr>
<tr>
<td>2</td>
<td>AE</td>
<td>AEAUTFD</td>
<td>Domain AE, variable AEAUTFD does not have the same length</td>
<td>char</td>
</tr>
<tr>
<td>3</td>
<td>CM</td>
<td>CMGRPID</td>
<td>Domain CM, variable CMGRPID is in cdpt9999/xx11111 but not in cdpt9999/xx22222</td>
<td>.</td>
</tr>
<tr>
<td>4</td>
<td>ZD</td>
<td>VISIT</td>
<td>Domain ZD, variable VISIT does not have the same format</td>
<td>char</td>
</tr>
<tr>
<td>5</td>
<td>ZD</td>
<td>ZDSPID</td>
<td>Domain ZD, variable ZDSPID does not have the same label</td>
<td>char</td>
</tr>
</tbody>
</table>

**Table 2**

<table>
<thead>
<tr>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Column Label</td>
<td>Column Format</td>
<td>Column Informat</td>
<td>Column Type</td>
<td>Column Length</td>
<td>Column Label</td>
<td>Column Format</td>
</tr>
<tr>
<td>1</td>
<td>Action taken with XXXXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Autopsy Findings</td>
<td>char</td>
<td>100</td>
<td>Autopsy Findings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Visit Name</td>
<td>char</td>
<td>200</td>
<td>Visit Name</td>
<td></td>
<td>$16.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Sponsor ID</td>
<td>char</td>
<td>200</td>
<td>Sponsor-Defined Identifier</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Automated Checking Tool

VALUE CHECK

• Checks differences in the controlled terminology
• Values checked only on common variables on common domains
• List of variables to be excluded added to the macro
Automated Checking Tool

VALUE CHECK

• Current tool runs better on events and intervention domains
  → all observations considered for each variable

• Findings domains are more problematic
  → checks not performed on a parameter level

• Possible future enhancement of the tool.
Tool Benefits

✓ Time/cost saving over manual checks

✓ Datasets more consistent $\rightarrow$ Improved pooling potential

✓ Run at start of a study $\rightarrow$ Pooling considered early
  $\rightarrow$ No impact on analysis programs

✓ Run of an individual study $\rightarrow$ monitor mapping/data changes

✓ Development of standard programs and maintenance $\rightarrow$ targeted
Tool Benefits

LB.LBMETHOD:

STUDY 1
PHOTOGRAPH

STUDY 2
PHOTOGRAPH

Standard/Project Program extract:

```plaintext
data example;
  set dataset1;
  if LBMETHOD = 'PHOTOGRAPH' then do;
    ...
  end;
run;
```

STUDY 3

LB.LBMETHOD:
PHOTOGRAPH

Known new term in LBMETHOD → Search for programs containing LBMETHOD
Tool Benefits

- **Time/cost saving** over manual checks
- **Datasets more consistent** → Improved pooling potential
- **Run at start of a study** → Pooling considered early
  → No impact on analysis programs
- **Run of an individual study** → monitor mapping/data changes
- **Development of standard programs** and maintenance → targeted
- **Extends to Analysis Datasets** or SDTMs constructed on different versions of the model
Conclusion

- Plan for pooling activities early
- More consistent SDTMs → better pooling & standard programs
- Develop an automated tool
- Further benefits to this tool than just SDTM consistency
- *Why* do these differences exist?
- Never assume SDTMs will be structurally identical
Doing now what patients need next