CDISC Italian User Network 2021

Virtual Event | 3 December 2021
ADaM traceability in a respiratory trial

Paola Vaghi
Head of Statistical Programming
Chiesi Farmaceutici S.p.A.

3 December 2021
Agenda

- Why traceability
- Reference Guidelines
- When traceability
- Traceability examples
- Take home messages
Why traceability

- An important component of regulatory review is an understanding of the provenance of the data
- Traceability permits an understanding of the relationships between the analysis results (tables, listings and figures in the study report), analysis datasets, tabulation datasets, and source data
- If the reviewer is unable to trace study data from data collection to the analyses, the regulatory review of a submission may be compromised
Respiratory trials Traceability – Reference Guidelines

**CDISC ADaM Guidelines**
- ADaM IG v1.1 package
- ADaM IG v1.2 package can be followed for new studies

**CDISC TAUG**
- COPD TAUG v1.0
- Asthma TAUG v1.0

**Regulatory Authorities**
- FDA Study Data Technical Conformance Guide v4.8.1 (October 2021)
- China Guideline on the Submission of Clinical Trial Data (July 2020)
- Pmda Revision of Basic Principles on Electronic Submission of Study Data for New Drug Applications (March 18, 2020)
Traceability of the results back to CRF data

Study documents:
- Statistical Analysis Plan
- Data Review Report

Data:
- CRF
- SDTM
- ADaM
- TLFs

Traceability documents:
- aCRF
- define.xml, SDRG
- Data point traceability
- Metadata traceability (define.xml, ADRG)
- Analysis Result Metadata

ADaM Traceability in a Respiratory Trial
Traceability in ADaM—When?

- Key dates
- When data are excluded from the analyses
  - According to rules defined in the SAP
  - Due to deviations identified during the Data Review Meeting
- On the efficacy parameters, especially when
  - the efficacy assessments are to be derived from observed data
  - Multiple merging of SDTM datasets is required
  - In case of estimands
- When the analysis is focused on a selection of cases (for example a subset of TEAEs of special interest)
Traceability in ADSL

- Common ADSL variables are copied from SDTM
- Other ADSL variables are derived within the ADSL dataset

For many ADSL variables metadata traceability is sufficient

Metadata traceability should clearly identify the source variables and should describe the steps followed to populate the variable
Traceability in ADSL– An Example on key dates

Pay attention: Unclear specifications are not good traceability

<table>
<thead>
<tr>
<th>Variable</th>
<th>Label</th>
<th>Origin / Source / Method / Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>RANDDT</td>
<td>Date of Randomization</td>
<td>Derived: Randomisation date</td>
</tr>
</tbody>
</table>

Source variables should be clearly identified

<table>
<thead>
<tr>
<th>Variable</th>
<th>Label</th>
<th>Origin / Source / Method / Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>RANDDT</td>
<td>Date of Randomization</td>
<td>Derived: Numeric version of DS.DSSTDTC when DS.DSDECOD=&quot;RANDOMIZED&quot;. If any part of the date is missing do not impute</td>
</tr>
</tbody>
</table>

Copy&paste of the SAS code should be avoided

```
data rand;
set sdtm.ds;
where dsdecod eq 'RANDOMIZED';
if length(dsstdtc)=10 then
    randdt=input(dsstdtc,is8601da.);
else if length(dsstdtc) gt 10 then
    randdt=datepart(input(dsstdtc,is8601dt.));
run;
```

Provide the source and detailed steps involved in the variable derivation (english language)
Traceability in efficacy endpoints (BDS structure)

**Efficacy Variable:** Pre-dose morning FEV1 values at each visit will be summarised by treatment group using descriptive statistics.

**Data point traceability:** each BDS record can retain SDTM variables to identify the source SDTM records and to help verifying records.

<table>
<thead>
<tr>
<th>AVISIT</th>
<th>PARAMCD</th>
<th>AVAL</th>
<th>BASE</th>
<th>CHG</th>
<th>RESTRESN</th>
<th>RESEQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visit 1 (Week -4)</td>
<td>FEV1</td>
<td>0.739</td>
<td>0.541</td>
<td>0.198</td>
<td>0.739</td>
<td>1</td>
</tr>
<tr>
<td>Visit 1 (Week -4)</td>
<td>FEV1</td>
<td>0.834</td>
<td>0.541</td>
<td>0.293</td>
<td>0.834</td>
<td>2</td>
</tr>
<tr>
<td>Visit 2 (Week 0)</td>
<td>FEV1</td>
<td>0.541</td>
<td>0.541</td>
<td>0</td>
<td>0.541</td>
<td>3</td>
</tr>
<tr>
<td>Visit 3 (Week 4)</td>
<td>FEV1</td>
<td>0.617</td>
<td>0.541</td>
<td>0.076</td>
<td>0.617</td>
<td>4</td>
</tr>
<tr>
<td>Visit 4 (Week 12)</td>
<td>FEV1</td>
<td>0.619</td>
<td>0.541</td>
<td>0.078</td>
<td>0.619</td>
<td>5</td>
</tr>
<tr>
<td>Visit 5 (Week 18)</td>
<td>FEV1</td>
<td>0.604</td>
<td>0.541</td>
<td>0.063</td>
<td>0.604</td>
<td>6</td>
</tr>
<tr>
<td>Visit 6 (Week 24)</td>
<td>FEV1</td>
<td>0.538</td>
<td>0.541</td>
<td>0.003</td>
<td>0.538</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RESTRESN</th>
<th>Numeric Result/Findings in Standard Units</th>
<th>float</th>
<th>REST12</th>
<th>Predecessor: RE,RESTRESN</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTRGR</td>
<td>BTR Grade after OverRead</td>
<td>text</td>
<td>13</td>
<td>Predecessor: RE,BTRGR</td>
</tr>
<tr>
<td>RESEQ</td>
<td>Sequence Number</td>
<td>integer</td>
<td>8</td>
<td>Predecessor: RE,RESEQ</td>
</tr>
</tbody>
</table>

**ADaM Traceability in a Respiratory Trial**
Traceability when multiple datasets are merged

**Efficacy Endpoint: FEV1 response at Visit 6:**
- Responder: change from baseline in pre-dose morning FEV1 >=100mL;
- Non-responder: change from baseline in pre-dose morning FEV1 <100mL;
- Non-responder: missing FEV1 at Visit 6

**Data point traceability:** as ADaM datasets are used as input for another ADaM dataset SRCDOM, SRCVAR, SRCSEQ should be used to guarantee traceability

<table>
<thead>
<tr>
<th>SUBJID</th>
<th>AVALC</th>
<th>AVALCAT1</th>
<th>SRCDOM</th>
<th>SRCVAR</th>
<th>SRCSEQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>Non-Responder</td>
<td>Non-Responder</td>
<td>ADRE</td>
<td>CHGCAT1</td>
<td>14</td>
</tr>
<tr>
<td>2003</td>
<td>Non-Responder</td>
<td>Non-Responder</td>
<td>ADRE</td>
<td>CHGCAT1</td>
<td>14</td>
</tr>
<tr>
<td>2005</td>
<td>Non-Responder</td>
<td>Non-Responder</td>
<td>ADRE</td>
<td>CHGCAT1</td>
<td>14</td>
</tr>
<tr>
<td>2006</td>
<td>Responder</td>
<td>Responder</td>
<td>ADRE</td>
<td>CHGCAT1</td>
<td>18</td>
</tr>
<tr>
<td>2007</td>
<td>Missing</td>
<td>Non-Responder</td>
<td>ADSL</td>
<td>CHGCAT1</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>Missing</td>
<td>Non-Responder</td>
<td>ADSL</td>
<td>CHGCAT1</td>
<td></td>
</tr>
</tbody>
</table>
Traceability when using a look-up table

- **Class-related TEAEs**: The number of class-related TEAEs and the number and percentage of subjects who experienced at least one class-related TEAE will be presented by treatment.

- A look-up table is created with the list of Class Related AEDECOD avoiding a long list of «if-then-else» programming statement.

- The look-up table is mentioned in the `define.xml`

| CQ01NAM | Populated by merging SDTM.AE with the look-up table LUCRTEAE. Set to "CLASS RELATED TEAE" if the preferred term in one of those listed in the look-up table; null otherwise |

- The look-up table is described in ADRG

1.4 Source Data Used for Analysis Dataset Creation
The ADaM datasets were derived from SDTM version 1.4. The datasets were derived from the final locked database.

In addition to the clinical database, the source data include:
- a lookup file that was used to include the medication common names associated with medication preferred names to be used in the tables,
- a lookup file that was used to identify patients with a pacemaker or with atrial fibrillation.
- a lookup file that was used to calculate SCRQ total and domain scores.

Details about these lookup tables are in the Appendix. The lookup table spreadsheets were converted to SAS transport files and included with the ADaM datasets.

- Following FDA guidance, the look-up table is submitted as a SAS transport file within the submission package.
Traceability in case of timepoints exclusion from the analyses

In documents

Efficacy: change from baseline in FEV1 AUC 0-12h

- The following timepoints should be excluded from SAP:
  - in case of more than two missing consecutive timepoints the AUC should be missing
  - In case of more than three timepoints the AUC should be missing
  - etc

- From Data Review Report:
  - from the PP analysis timepoints collected after 6h from rescue intake will be excluded

In ADaM

- An algorithm is applied to exclude records as per SAP rule
- Timepoints to be excluded as per DRR are mapped into SDTM.DV
- To improve traceability a look-up table can be created from the table in DRR to clearly identify timepoints to be excluded
- SDTM.RE, SDTM.DV and the look-up table are joined to identify both
  - Timepoints to be included in AUC derivations
  - AUC that cannot be derived for more than allowed missing timepoints
Complex traceability example

- In case the parameter is heavily derived from collected data, multiple approaches should be put in place to guarantee the traceability

Efficacy endpoint in respiratory trials: the rate of Moderate and Severe exacerbations

- Exacerbation data as in the eCRF are far away from exacerbations variables used in the analyses
- Exacerbations in SDTM.AE are combined with multiple SDTMs with related informations -> ADEXAC
- The rate of moderate and severe exacerbations is stored in ADXASUM -> ADXASUM
Complex traceability example – Respiratory exacerbations

Traceability using and intermediate dataset

Example: The number and percentage of exacerbations treated with systemic corticosteroids and antibiotics are presented by treatment group

- The intermediate dataset ADCM is derived to clearly identify records taken from SDTM.CM that should be joined to SDTM.AE

ADCM (OCCDS structure)
- Analysis flags to identify:
  - all the medications related to an exacerbation (ANL01FL)
  - systemic corticosteroids for an exacerbation (ANL02FL)
  - antibiotics for an exacerbation (ANL03FL)

ADEXAC (OCCDS structure)
- ESATRT is populated from the analyses flags derived in ADCM

<table>
<thead>
<tr>
<th>SUBID</th>
<th>ASEQ</th>
<th>AETERM</th>
<th>ESATRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>1</td>
<td>COPD Exacerbation</td>
<td>Systemic Corticosteroids and Antibiotics</td>
</tr>
<tr>
<td>2002</td>
<td>2</td>
<td>COPD Exacerbation</td>
<td>Systemic Corticosteroids and Antibiotics</td>
</tr>
<tr>
<td>2003</td>
<td>1</td>
<td>COPD Exacerbation</td>
<td>Systemic Corticosteroids and Antibiotics</td>
</tr>
<tr>
<td>2003</td>
<td>2</td>
<td>COPD Exacerbation</td>
<td>Systemic Corticosteroids and Antibiotics</td>
</tr>
<tr>
<td>2007</td>
<td>2</td>
<td>COPD Exacerbation</td>
<td>Antibiotics Only</td>
</tr>
</tbody>
</table>
Complex traceability example – Respiratory exacerbations

Traceability in combined episodes

Two COPD exacerbations will be considered as a single episode in the statistical analysis if:

- the second exacerbation started less than 10 days after the end of the systemic corticosteroids and/or antibiotics intake for the previous exacerbation
- the second exacerbation started less than 10 days after the onset of the previous exacerbation

All exacerbations are collected in ADaM.ADEXAC

EGRPID groups exacerbations that should be a single episode

An analysis flag on the first occurrence within a combined episode allows the right count of exacerbations
Complex traceability example – Respiratory exacerbations

Traceability with estimands

- Primary analysis- treatment policy strategy: n. of moderate and severe exacerbations occurring during the planned weeks of treatment
- Alternative estimand – hypothetical estimand: n. of moderate and severe exacerbations based only on treatment data

- **ADEXAC (OCCDS)**
  Two flags to identify exacerbation of each estimand
  - **ANL01FL** Flag to identify exacerbation occurred during the overall period: [...] 
  - **ANL03FL** Flag to identify exacerbation occurred during the on-treatment period: [...] 

- **ADXASUM (BDS)**
  Two different parameters for the count of moderate and severe exacerbation of each estimand

<table>
<thead>
<tr>
<th>SUBJID</th>
<th>PARAM</th>
<th>PARAMCD</th>
<th>PARAMNM</th>
<th>AVAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>002</td>
<td>Number of Moderate/Severe Exa (Hypothetical estimand)</td>
<td>MSEXH-E</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>002</td>
<td>Num of Moderate/Severe Exa (Treatment Policy)</td>
<td>MSEXTP</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Traceability in ADRG

A description of the dataset in ADRG can support in complex data derivations

- Compliance is heavily derived from data collected in the daily diary and the intake at the clinic
- A description on how daily intake have been considered for the compliance purposes compared to how they are collected in the diary is provided in ADRG

Analysis outputs programs can be submitted helping the reviewer in understanding the process by which the variables for the analyses were derived

7. Submission of Programs

All programs for analysis datasets and primary and secondary efficacy results are submitted. They were all created on a SAS platform using version 9.4. The internal reference date used to create dates in ADaM datasets is January 1, 1900.

7.1 ADaM Programs

<table>
<thead>
<tr>
<th>Program Name</th>
<th>Output</th>
<th>Macro Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adit.sas</td>
<td>ADITT</td>
<td>Clearwork, adl_trtdate, date, mrgesupp, adls trope, adls cas, adls phase, period, adlcheck, adlancs, adlance, adlspec</td>
</tr>
<tr>
<td>Adls.sas</td>
<td>ADSL</td>
<td>Clearwork, date, mrgesupp, adls trope, adls phase, period, adltrate, adls smoking, adltrimht, adlcheck, adlancs, adlspec</td>
</tr>
</tbody>
</table>

7.2 Analysis Output Programs

<table>
<thead>
<tr>
<th>Program Name</th>
<th>Output Number</th>
<th>Title</th>
<th>Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>t_re_fev1_moe_model_it.sas</td>
<td>14.2.1.1.2</td>
<td>Table: Statistical Analysis of Change from Baseline in Pre-dose Morning FEV1 (L) (Intention-to-Treat set)</td>
<td>ADRE</td>
</tr>
<tr>
<td>t_re_fev1_moe_model_pp.sas</td>
<td>14.2.1.2.2</td>
<td>Table: Statistical Analysis of Change from Baseline in Pre-dose Morning FEV1 (L) (Per Protocol set)</td>
<td>ADRE</td>
</tr>
<tr>
<td>t_re_fev1resp_model_it.sas</td>
<td>14.2.2.2.2</td>
<td>Table: Statistical Analysis of FEV1 Response at Visit 6 (Week 24) (Intention-to-Treat set)</td>
<td>ADRESUM</td>
</tr>
</tbody>
</table>
Take home messages

• Traceability is an important component in data submission; poor traceability can compromise the entire submission

• Data point traceability should be applied if feasible as it’s the best way to identify the source records

• Use of an intermediate datasets, intermediate parameters and look-up tables can improve traceability. Data dependency and intermediate datasets should be described in the ADRGs; a graphical representation is better

• Metadata traceability should clearly identify the source data and should describe all the steps to be followed to populate derived data

• A description of the datasets in the ADRG can support complex data derivations

• Analysis output programs can be submitted helping the reviewer in understanding the process by which the variables were derived
Thank You

Paola Vaghi
P.vaghi@chiesi.com
Back-up slides
Traceability when adding a Row to a BDS structure

Safety: in case of multiple measurements associated to the same timepoint the average value will be considered for SBP and BDP

- **Data point traceability:** DTYPE variable is used to indicate when a new derived row has been added to the dataset and to define how the analysis values was derived

<table>
<thead>
<tr>
<th>SUBJID</th>
<th>AVISIT</th>
<th>AVSITN</th>
<th>ADT</th>
<th>ATM</th>
<th>ADTM</th>
<th>PARAMCD</th>
<th>PARAMN</th>
<th>AVAL</th>
<th>DTYPE</th>
<th>ANL01FL</th>
<th>ANL02FL</th>
<th>VISITNUM</th>
<th>VSSEQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>3006</td>
<td>Vis1</td>
<td>1</td>
<td>09OCT2019</td>
<td>9:30</td>
<td>09OCT19:09:30:00</td>
<td>SYSBP</td>
<td>1</td>
<td>156</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>3006</td>
<td>Vis1</td>
<td>2</td>
<td>04NOV2019</td>
<td>8:42</td>
<td>04NOV19:08:42:00</td>
<td>SYSBP</td>
<td>1</td>
<td>153</td>
<td>Y</td>
<td></td>
<td></td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>3006</td>
<td>Vis3</td>
<td>3</td>
<td>04DEC2019</td>
<td>8:30</td>
<td>04DEC19:08:30:00</td>
<td>SYSBP</td>
<td>1</td>
<td>146</td>
<td>Y</td>
<td></td>
<td></td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>3006</td>
<td>Vis5</td>
<td>5</td>
<td>12MAR2020</td>
<td>9:00</td>
<td>12MAR20:09:00:00</td>
<td>SYSBP</td>
<td>1</td>
<td>110</td>
<td>Y</td>
<td></td>
<td></td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>3006</td>
<td>Vis5</td>
<td>5</td>
<td>12MAR2020</td>
<td>9:24</td>
<td>12MAR20:09:24:00</td>
<td>SYSBP</td>
<td>1</td>
<td>174</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>3006</td>
<td>Vis5</td>
<td>5</td>
<td>12MAR2020</td>
<td>12:20</td>
<td>12MAR20:12:20:00</td>
<td>SYSBP</td>
<td>1</td>
<td>173.5</td>
<td>AVERAGE</td>
<td>Y</td>
<td></td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

DTYPE=AVERAGE when AVAL is the average of the two-pre-dose assessments
ANL01FL=Y only on records to be analysed
VSSEQ is missing for the average record
The change from baseline of pre-dose FEV1 was analyzed by means of a MMRM with treat, visit, treat by visit interaction, [...] and base by visit interaction as cov.