

# Let Us Trace Back to the Beginning from the End

Madhusudhan Reddy Papasani, Merck & Co., Inc., Upper Gwynedd, PA  
 Swathi Kotla, Merck & Co., Inc., Upper Gwynedd, PA  
 Elisabeth Pyle, Merck & Co., Inc., Upper Gwynedd, PA

## ABSTRACT:

Integrity of regulatory submissions relies on traceability, allowing regulators to check source data and understand the processes involved in the transformation of data, to the analysis results. CDISC standards, along with the Define-XML, are interconnected and designed to provide relationships between analysis results, analysis data and SDTM datasets. In this paper, a discussion of data standards and traceability is reviewed using an analysis table from an oncology Overall Survival (OS) analysis to explore traceability from the analysis results back to the case report form (CRF) collected data. As part of this process, interrelationships between analysis results (AR) to analysis results metadata (ARM), ARM to ADaM datasets, ADaM datasets to SDTM datasets and finally SDTM datasets to the CRF will be presented, traced and highlighted here in the same way a regulatory reviewer would trace results to the source data in a submission review.

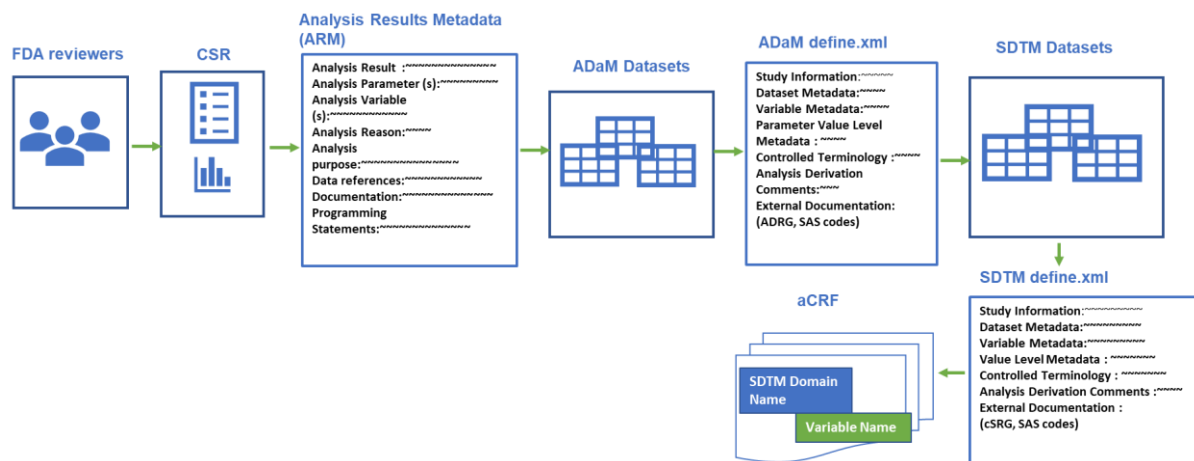
## INTRODUCTION:

CDISC Foundational Standards and Therapeutic Area Standards in the form of models, Implementation Guides, and Therapeutic Area User Guides serve as useful resources in the design of clinical study data. In addition, CDISC has developed the Define-XML to represent metadata for datasets, and then published the Analysis Results Metadata (ARM) as an extension to the Define-XML for table and figures metadata. The robust repository of CDISC standards and specifications for data representation and their suitability to various clinical situations led to their implementation for regulatory submission of clinical trial data and reporting. The FDA endorsed CDISC data standards and expects sponsors to implement CDISC standards for any studies that have started after Dec 17, 2016.

CDISC Foundational Standards are intended to work together and are inter-related so that one data standard flows into the next, enhancing the quality of data from beginning to end. Three of these data standards are: CDASH standards for data collection and case report form (CRF) development, the SDTM standards for standardizing collected data, and the ADaM standards for analysis datasets and they work in tandem to support the collection and define the structure of clinical trial data through statistical analyses. The design of datasets (SDTM, ADaM) and their associated metadata (define.xml) including the recently added ARM should establish unequivocal communication in the form of traceability from beginning to end and vice versa.

As shown in Fig. 1, the Analysis Results Metadata provides information about the purpose of the analysis, input ADaM datasets and analysis methods with reference to the clinical study report (CSR). The Define.xml for ADaM datasets provides the structure and content of ADaM datasets and documentation of input SDTM datasets, while the define.xml for SDTM datasets assures data tabulation integrity referring to CRF annotations.

Fig. 1. The relationships of CDISC standards in the FDA review process starting from CSR or “the end”.



In oncology trials, overall survival (OS) is a key efficacy endpoint. Clinical trials are designed to test if an experimental drug prolongs OS compared to existing therapies. For illustration purposes, a Therapy per Physician's Choice (TPC), which defines existing disease primary therapies, will be compared with a new therapy targeting a biomarker (BM) or a Therapy Targeting Biomarker (TTB). This paper describes how to review OS analysis tables and trace back to CRF collected data using CDISC Foundational Standards.

## OVERALL SURVIVAL(OS):

Overall survival is defined as the time from randomization to death due to any cause. Subjects without documented death at the time of the analysis are censored at the last known alive date. When the analysis cutoff date is implemented, only deaths before the cutoff date are considered as events. The subjects are censored at the cutoff date, if the subject either lived or died after the cutoff date. Overall Survival typically includes Kaplan-Meier (KM) estimates of survival (median and its confidence intervals). Additionally, the two treatments groups are compared by computing hazard ratios and p-values. A p-value of less than 0.05 indicates a significant treatment effect.

## THE STUDY DESIGN:

|                                               |                                                                                                                                                                                                                                                                                                               |
|-----------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Study Design Overview                         | A Phase III Randomized Trial of Therapy Targeting Biomarker (TTB) versus therapy per physician's choice (TPC) for Metastatic Cancer.                                                                                                                                                                          |
| Treatment Assignment                          | Subjects were randomized in a 1:1 ratio between the two treatment groups: (1) TTB arm and (2) TPC arm. Stratification factor : BM status (positive [BM+] vs negative [BM-])                                                                                                                                   |
| Analysis Populations                          | Efficacy: Intention-to-treat Population (ITT)                                                                                                                                                                                                                                                                 |
| Primary Endpoint(s)                           | OS in all subjects                                                                                                                                                                                                                                                                                            |
| Statistical Methods for Key Efficacy Analyses | The non-parametric Kaplan-Meier method will be used to estimate the median survival rates within each treatment group.<br>Estimation of the hazard ratio will be performed using a stratified Cox regression model.<br>The treatment difference in survival will be assessed by the stratified log-rank test. |

**Table 1 Example CSR table summarizing OS**

## Analysis of Overall Survival All Subjects (ITT Population)

|                                                                      | TTB<br>(N = ###)                 | TPC<br>(N = ###)     |
|----------------------------------------------------------------------|----------------------------------|----------------------|
| Kaplan-Meier Estimates (Months) <sup>†</sup><br>Median (95% CI)      | ##. # (##. #, ##. #)             | ##. # (##. #, ##. #) |
| vs TPC<br>Hazard Ratio (95% CI) <sup>‡</sup><br>p-value <sup>§</sup> | #. ## (#. ##, #. ##)<br>0. ##### |                      |

<sup>†</sup> From product-limit (Kaplan-Meier) method for censored data.

<sup>‡</sup> Based on Cox regression model with Efron's method of tie handling with treatment as a covariate and stratified by BM status (BM+ vs BM-).

<sup>§</sup> p-value based on stratified log-rank test.

## QUICK VALIDATION OF RESULTS USING ARM:

In a review of this CSR, the primary hypothesis of OS is the key objective of the study and the OS table results and related p-values produced by the stratified log-rank test follow as the result of interest for the reviewer. The analysis result metadata (Table 2) is submitted and the ARM provides:1.) a cross-reference for the result of interest within table display to the analysis dataset(s), parameter(s) and analysis variables along with the sub-setting conditions, 2.) a snippet of programming code for a quick validation of the results 3.) details for the primary endpoint results and 4.) information on the analysis purpose, the analysis reason, the statistical methodology and references to external files such as the statistical analysis plan (SAP) or SAS® programs etc.

As noted below in Table 2, the ARM identifies ADTTE as the input dataset for the OS analysis with the subset condition of ITTFL="Y". The parameter = "Overall Survival" with PARAMCD= "OS", and the analysis variables: AVAL, CNSR, STRATAR, ITTFL are used in the statistical model to generate p-values using the stratified log-rank test. The programming code snippet provides syntax for implementing the proc life test with stratified log-rank test to generate the p-value. This allows

the reviewer to follow the results from the CSR to the data quickly and in turn, validate the p-value. In this example, a clear metadata traceability has been established between the p-value (the treatment difference in survival assessed by the stratified log-rank test) and the analysis dataset: ADTTE and its parameters or variables along with SAS procedures.

**Table 2 Analysis Results Metadata**

|                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|--------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Display                                    | Table 1. Analysis of Overall Survival in all Subjects (ITT Population)                                                                                                                                                                                                                                                                                                                                                                     |
| Analysis Parameter(s)                      | ADTTE.PARAMCD= 'OS'                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Analysis Result                            | The treatment difference in survival assessed by the stratified log-rank test                                                                                                                                                                                                                                                                                                                                                              |
| Analysis Variable(s)                       | ADTTE.AVAL, ADTTE.CNSR                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Analysis Reason                            | Efficacy Analysis                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Analysis Purpose                           | Primary Endpoint: Overall Survival                                                                                                                                                                                                                                                                                                                                                                                                         |
| Data References (incl. Selection Criteria) | ADTTE. ITTFL='Y'                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Documentation                              | TTP is compared to TPC on OS using a stratified Log-rank test. Please refer to SAP section.8 for the Statistical Methods for Efficacy Analyses                                                                                                                                                                                                                                                                                             |
| Programming Statements                     | <p><b>p-values:</b><br/> <b>proc lifetest</b> data = ADTTE;<br/>             where TRT01PN in (1,2);<br/>             time AVAL*CNSR(1);<br/>             strata STRATAR/group=TRT01PN;<br/>             ods output HomTests=pval(where=(test='Log-Rank'));</p> <p><b>run;</b><br/>         Further reference: by checking program:<br/>         SAS Version: [SAS Version 9. 4]<br/>         SAS program: e00r0anal0tte0ver0osgen.txt</p> |

## REVIEW OF ADTTE (ADaM TIME-TO-EVENT ANALYSIS DATASET):

Once the analysis dataset(s) and its parameters or variables used to generate the analysis results are identified, then it is important for the reviewer to determine the validity of the analysis data by tracing the dataset contents and derivations. In this case, it is important to determine if PARAMCD="OS" and analysis variables (AVAL, CNSR and STRATAR) are properly populated in the analysis dataset: ADTTE. To follow the derivations, there are multiple steps and multiple datasets that require referencing as outlined in metadata for ADTTE.

Table 4 provides the data snippet of ADTTE with the data elements used for the analysis and Table 5 presents the metadata describing how the content of ADTTE is populated. Upon review, it is clear ADT is an important variable that has been derived from ADT of the intermediate dataset ADINTDT. The metadata describes how ADTTE.ADT is selected from ADINTDT. On the other hand, ADTTE.STARTDT has a predecessor as ADSL.RANDDT. The AVAL variable is derived using ADTTE.ADT and ADTTE.STARTDT. The CNSR variable describes if the subject had an event of interest (CNSR=0) or not (CNSR=1). Lastly, ADTTE.STRATAR, ADTTE.ITTFL are copied from ADSL as is.

To further understand how ADINTDT is processed to create ADTTE, the reviewer may refer to the ADTTE programming code. ADINTDT has all the different dates for both event or censor and these are presented using a BDS dataset. ADINTDT.PARAMCD, ADINTDT.ADT and ADINTDT.ASEQ variables are used to create ADTTE dataset. The dataset ADINTDT is in a vertical structure and in order to use it for ADTTE, it must be transposed to horizontal structure for the dates and ASEQ. The programming code to transpose the data is displayed below and the resulting temporary dataset (\_\_adintdt2) is partially displayed in Table 3.

|                                                                                                                                                                |                                                                                                                                                                                                   |                                                                                                     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|
| <pre>Dates of interest in a single row for the subject proc transpose data=adintdt out=dates(drop=_name_ _label_); by usubjid; id paramcd; var adt; run;</pre> | <pre>*ASEQ numbers for the dates of interest in a single row for the subject; proc transpose data= adintdt out=seq(drop=_name_ _label_) prefix=seq_; by usubjid; id paramcd; var aseq; run;</pre> | <pre>*Merging dates and sequence datasets; data __adintdt2; merge dates seq; by usubjid; run;</pre> |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|

**Table 3 Transposed ADINTD as a temporary SAS dataset \_\_adintdt2**

| USUBJID     | CUTOFFDT   | DTHDT      | LKALDT     | UNCUTDT    | SEQ_CUTOFFDT | SEQ_DTHDT | SEQ_LKALDT | SEQ_UNCUTDT |
|-------------|------------|------------|------------|------------|--------------|-----------|------------|-------------|
| XYZ-001-001 | 02-15-2017 |            | 02-01-2017 |            | 1            |           | 3          |             |
| XYZ-001-002 | 02-15-2017 | 04-14-2016 | 04-12-2016 |            | 1            | 2         | 3          |             |
| XYZ-001-003 | 02-15-2017 |            | 02-05-2017 | 02-20-2017 | 1            |           | 3          | 4           |
| XYZ-001-004 | 02-15-2017 | 08-12-2016 | 08-10-2016 |            | 1            | 2         | 3          |             |

This temporary dataset is optimal for deriving PARAM, PARAMCD, ADT, AVAL, and CNSR. The variables: SRCDOM, SRCVAR have been assigned as 'ADINTDT', and 'ADT' respectively. SRCSEQ contains the value of ASEQ on the record in ADINTDT from which ADTTE.ADT has been selected. The following code is helpful for the reviewer to check data in ADTTE and to determine whether it is in alignment with the metadata/programming code.

```

Deriving/Assigning PARAM, PARAMCD, ADT, AVAL, CNSR, SRCDOM, SRCVAR and SRCSEQ;
data os(keep=usubjid paramcd startdt adt cnsr evntdesc srcseq);
  set __adintdt2;
  length evntdesc $100 paramcd $8;
  srcdom= "ADINTDT";
  srcvar="ADT";
  startdt=randdt;
  paramcd="OS";
  if dthdt ne . then do;
    adt=dthdt;
    cnsr=0;
    evntdesc="DEATH";
    srcseq=seq_dthdt;
  end;
  else if uncutdt ne . then do;
    adt=cutoffdt;
    cnsr=1;
    evntdesc='CENSORED AT DATA CUTOFF DATE';
    srcseq=seq_cutoffdt;
  end;
  else if LKALDT ne . then do;
    adt=lkaldt;
    cnsr=1;
    evntdesc="LAST KNOWN ALIVE BEFORE CUTOFF";
    srcseq=seq_lkaldt;
  end;
  aval=adt-startdt+1;
  format startdt yymmdd10. ;
run;

```

The resulting ADTTE analysis dataset along with metadata is presented in Tables 4 and 5.

**Table 4 ADTTE Data**

| USUBJID     | PARAMCD | STARTDT    | ADT        | AVAL | SRCDOM  | SRCVAR | SRCSEQ | CNSR | STRATAR | ITTFI | TRT10P | TRT10PN |
|-------------|---------|------------|------------|------|---------|--------|--------|------|---------|-------|--------|---------|
| XYZ-001-001 | OS      | 06-11-2016 | 02-01-2016 | 246  | ADINTDT | ADT    | 3      | 1    | BM+     | Y     | TTB    | 1       |
| XYZ-001-002 | OS      | 02-28-2016 | 04-14-2016 | 47   | ADINTDT | ADT    | 2      | 0    | BM-     | Y     | TTB    | 1       |
| XYZ-001-003 | OS      | 01-04-2016 | 02-15-2016 | 408  | ADINTDT | ADT    | 1      | 1    | BM+     | Y     | TPC    | 2       |
| XYZ-001-004 | OS      | 05-25-2016 | 08-12-2016 | 80   | ADINTDT | ADT    | 2      | 0    | BM-     | Y     | TPC    | 2       |

**Table 5 ADTTE Variable Metadata**

| Variable Name | Variable Label                        | Type | Codelist/ Controlled Terms | Notes                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|---------------|---------------------------------------|------|----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| USUBJID       | Unique Subject Identifier             | Char |                            | ADSL.USUBJID                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| PARAMCD       | Parameter Code                        | Char | OS                         | For Overall Survival, set to 'OS'                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| STARTDT       | Time to Event Origin Date for Subject | Char |                            | ADSL.RANDDT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| ADT           | Analysis Date                         | Num  |                            | <p>Analysis date for the event (in this case Death) or for Censoring.<br/>If a subject died on or prior to the data cutoff date, then the analysis date is the date of death (Event).</p> <p><i>Programming logic:</i><br/>(If ADINTDT. ADT Where ADINTDT. PARAMCD= "DTHDT") is less than (ADINTDT. ADT where ADINTDT. PARAMCD= "CUTOFFDT") then set ADTTE.ADT to ADINTDT. ADT Where ADINTDT.PARAMCD= "DTHDT".<br/>Else if a subject is either alive or dead after cutoff date then the subject is censored on the cutoff date (Censor);<br/><i>Programming logic:</i><br/>Else if ADINTDT. ADT is not null where ADINTDT. PARAMCD= "UNCUTDT" then set ADTTE. ADT to ADINTDT. ADT Where ADINTDT. PARAMCD= "CUTOFFDT".<br/>Please note that uncut date is only populated for the subjects who have last known dates beyond cutoff date.</p> <p>Else if a subject is known alive prior to the cutoff date, and no available dates beyond cutoff, then the subject is censored at the last known alive date prior to cutoff (Censor).</p> <p><i>Programming logic:</i><br/>Else if ADINTDT.ADT is null where ADINTDT.PARAMCD= "UNCUTDT") and ADINTDT. ADT is not missing, where ADINTDT. PARAMCD="LKALDT ") then set ADTTE. ADT to ADINTDT. ADT Where ADINTDT. PARAMCD= " LKALDT ".</p> |
| AVAL          | Analysis Value                        | Num  |                            | <p>Time from randomization date to death date or censoring date.<br/><i>Programming logic:</i><br/>ADTTE. AVAL=ADTTE. ADT-ADTTE. STARTDT+1</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| SRCDOM        | Source Data                           | Char |                            | Assigned "ADINTDT"                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| SRCVAR        | Source Variable                       | Char |                            | Assigned "ADT"                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| SRCSEQ        | Source Sequence Number                | Num  |                            | ADINTDT.ASEQ of the observation in ADINTDT that satisfies the ADT derivation                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| CNSR          | Censor                                | Char |                            | 0 for event (Subject died on or prior to the data cutoff date); else 1 for censored subjects                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| STRATAR       | STRATA Used for Randomization         | Char |                            | ADSL.STRATAR                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| ITTFL         | Intent to Treat Population Flag       | Char |                            | ADSL.ITTFL                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| TRT01P        | Planned Treatment for Period 01       | Char |                            | ADSL.TRT01P                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| TRT01PN       | Planned Treatment for Period 01 (N)   | Num  |                            | ADSL.TRT01PN                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |

**REVIEW OF ADINTDT (INTERIM DATASET OF DATES USED FOR ADTTE):**

As discussed above, ADTTE.ADT is a critical variable and has its roots in ADINTDT dataset. ADINTDT is an intermediate dataset utilized to understand and trace the dates of interest which are important components of the derivations in ADTTE dataset. ADINTDT has all the dates of interest in BDS structure as parameters and ADTs. As such ADINTDT serves as bridge or intermediate dataset between SDTM datasets and ADTTE dataset. ADINTDT has the following parameters and these definitions are based on the hypothetical statistical analysis plan (as described in Metadata).

- Date of Data Cutoff: The cutoff date is a fixed date for the interim analysis and subjects are censored at the cutoff date if the last known contact date is beyond the cutoff date for the interim analysis.
- Date of Death: Captures death if it happened before the cutoff date and is considered an event.
- Date of Last Known Alive Prior to Cutoff: Captures the latest known alive date for the subject from SDTM domains including exposure, subject visits, tumor results, disease response, laboratory assessment, vital sign assessments, survival follow-up alive status, disposition, etc. before the cutoff date. The subjects who have no events and have adequate assessments before the Date of Data Cutoff are censored at the Date of Last Known Alive Prior to Cutoff.
- Date Beyond Cutoff: only appears if subjects have data available after the cutoff date.

ADINTDT is created by pre-processing SDTM or ADaM datasets. Each date of interest is derived and stored as a temporary dataset and all the datasets are ultimately set together. The reviewer may find it helpful to review the code along with the define.xml (Metadata) to understand how the parameters of interest are derived, preserving traceability to SDTM or other ADaM datasets using SRCDOM, SRCVAR and SRCSEQ

```
*Date of Data Cutoff;
%let cutoff_dt= 02-05-2017;
data cutoffdt(keep=usubjid paramcd adt);
  set adsl(keep=usubjid);
  length paramcd $8;
  paramcd="CUTOFFDT";
  param= "Date of Data Cutoff";
  paramn=1;
  adt=input("&cutoff_dt", ?? date9. );
  format adt yymmdd10. ;
run;
```

```
*Date of Death;
data dthdt;
  set dd;
  where ddgrpid='DEATH' and ddtstcd='PRCDTH' and
  dddtc ne '';
  adt=input(dddttc, yymmdd10. );
  srcdom= 'DD';
  srcvar= 'DDTTC';
  srcseq= DDSEQ;
  param='Date of Death';
  paramcd='DTHDT';
  paramn=2; if adt <="&cutoff_dt"d;
run;
```

```
*Date of Last Known Alive Prior to Cutoff;
data dts;
  set ex (in=xs keep=usubjid domain exstdtc exseq rename=(exstdtc=_dtc exseq=srcseq))
  ex(in=xr keep=usubjid domain exendtc exseq rename=(exendtc=_dtc exseq=srcseq))
  sv (in=sv keep=usubjid domain svstdtc rename=(svstdtc=_dtc))
  tr (in=tr keep=usubjid domain trdtc trseq rename=(trdtc=_dtc trseq=srcseq))
  rs (in=rs keep=usubjid domain rsdtc rsseq rename=(rsdtc=_dtc rsseq=srcseq))
  lb (in=lb keep=usubjid domain lbdtc lbseq rename=(lbdtc=_dtc lbseq=srcseq))
  vs(in=vs keep=usubjid domain vsdtc vsseq rename=(vsdtc=_dtc vsseq=srcseq))
  fa(in=fa keep=usubjid domain fadtc faseq fagrpid faorres rename=(fadtc=_dtc
  faseq=srcseq) where=(fagrpid='SFU' and upcase(faorres)='ALIVE'))
  ds(in=ds keep=usubjid domain dsdtc dsseq dsgrpid rename=(dsdtc=_dtc
  dsseq=srcseq) where=(dsgrpid='SR'))
  ss(in=ss keep=usubjid domain ssdtc ssseq rename=(ssdtc=_dtc ssseq=srcseq));
  if _dtc>' ' and input(_dtc, ?? yymmdd10. ) <="&cutoff_dt"d;
  length srcdom srcvar $8;
  srcdom=domain;
  if xs then srcvar='EXSTDTC';
  else if xr then srcvar='EXENDTC';
  else if sv then srcvar='SVSTDTC';
  else if tr then srcvar='TRDTC';
  else if rs then srcvar='RSDTC';
  else if lb then srcvar='LBDTC';
  else if vs then srcvar='VSDTC';
  else if fa then srcvar='FADTC';
  else if ds then srcvar='DSDTC';
  else if ss then srcvar='SSDTC';
  adt=input(_dtc, ?? yymmdd10. );
  format adt yymmdd10. ;run;
```

```
*Sort the resulting dataset and select the latest date and assign srcdom,srcvar;
proc sql;
  create table __dts_tmp as
  select d. * from dts as d
  order by usubjid, adt, srcdom desc, srcvar desc, srcseq;
quit;
data lnaltdt ;
set __dts_tmp;
  by usubjid adt descending srcdom descending srcvar srcseq;
param='Date of Last Known Alive Prior to Cutoff ' ;
paramcd='LKALDT';
paramn=3;
if last. usubjid;
format adt yymmdd10. ; run;
```

Similar to the Date of Last Known Alive Prior to Cutoff, the Date Beyond Cutoff is derived, but only populated for the subjects if the last known date is beyond the cutoff date. All the dates of interest are derived and kept as temporary datasets. Finally, set all the temporary datasets that have dates of interest to create the ADINTDT dataset. For illustration purposes, only four parameters are discussed, but any additional dates of interest are added, following the examples above, to create a comprehensive ADINTDT.

ASEQ the record identified in ADINTDT has the same value as PARAMN. The PARAMN variable is unique with a subject. SRCSEQ in ADTTE contains the same value of ASEQ of ADINTDT and is helpful in establishing traceability to the source record in ADINTDT from which ADTTE.ADT has been selected. On the other hand, the traceability triplet: SRCDOM, SRCSEQ, and SRCVAR in ADINTDT navigate the reviewer to SDTM records so the reviewer would be able to verify the data to the collected source data.

```
data adintdt;
set cutoffdt dthdt lnaladt uncutdt;
    aseq=paramn;
run;
```

**Table 6 ADINTDT Data**

| USUBJID     | PARAM                                    | PARAMCD  | PARAMN | ADT        | SRCDOM | SRCVAR | SRCSEQ | ASEQ |
|-------------|------------------------------------------|----------|--------|------------|--------|--------|--------|------|
| XYZ-001-001 | Date of Data Cutoff                      | CUTOFFDT | 1      | 02-15-2017 |        |        |        | 1    |
| XYZ-001-001 | Date of Last Known Alive Prior to Cutoff | LKALDT   | 3      | 02-01-2017 | FA     | FADTC  | 58     | 3    |
| XYZ-001-002 | Date of Data Cutoff                      | CUTOFFDT | 1      | 02-15-2017 |        |        |        | 1    |
| XYZ-001-002 | Date of Death                            | DTHDT    | 2      | 04-14-2016 | DD     | DDTDC  | 1      | 2    |
| XYZ-001-002 | Date of Last Known Alive Prior to Cutoff | LKALDT   | 3      | 04-12-2016 | LB     | LBDTC  | 1800   | 3    |
| XYZ-001-003 | Date of Data Cutoff                      | CUTOFFDT | 1      | 02-15-2017 |        |        |        | 1    |
| XYZ-001-003 | Date of Last Known Alive Prior to Cutoff | LKALDT   | 3      | 02-05-2017 | EG     | EGDTC  | 322    | 3    |
| XYZ-001-003 | Date Beyond Cutoff                       | UNCUTDT  | 4      | 02-20-2017 | RS     | RSDTC  | 161    | 4    |
| XYZ-001-004 | Date of Data Cutoff                      | CUTOFFDT | 1      | 02-15-2017 |        |        |        | 1    |
| XYZ-001-004 | Date of Death                            | DTHDT    | 2      | 08-12-2016 | DD     | DDTDC  | 1      | 2    |
| XYZ-001-004 | Date of Last Known Alive Prior to Cutoff | LKALDT   | 3      | 08-10-2016 | LB     | LBDTC  | 1322   | 3    |

**Table 7a ADINTDT Variable level Metadata**

| Variable Name | Variable Label            | Type | Codelist/Controlled Terms | Notes                                                                      |
|---------------|---------------------------|------|---------------------------|----------------------------------------------------------------------------|
| USUBJID       | Unique Subject Identifier | Char |                           | ADSL. USUBJID                                                              |
| PARAMCD       | Parameter Code            | Char |                           | Assigned based on Parameter Value "CUTOFFDT", "DTHDT", "LKALDT", "UNCUTDT" |

| Variable Name | Variable Label           | Type | Codelist/Controlled Terms | Notes                                                                                                                                                       |
|---------------|--------------------------|------|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PARAM         | Parameter                | Char |                           | Set ADINTDT. PARAM to the following Values:<br>Date of Data Cutoff,<br>Date of Death,<br>Date of Last Known Alive Before Data Cutoff,<br>Date Beyond Cutoff |
| PARAMN        | Parameter (N)            | Num  |                           | Assign Values as below for ADINTDT.PARAMCDs:<br>1 for CUTOFFDT,<br>2 for DTHDT,<br>3 for LKALDT,<br>4 for UNCUTDT.                                          |
| ADT           | Analysis Date            | Num  |                           | See parameter-level metadata                                                                                                                                |
| AVALC         | Analysis Value           | Num  |                           | Character version of ADINTDT ADT                                                                                                                            |
| SRCDOM        | Source Data              | Char |                           | Set to SDTM domain from which ADT is selected                                                                                                               |
| SRCVAR        | Source Variable          | Char |                           | Set to SDTM Variable from which ADT is selected                                                                                                             |
| SRCSEQ        | Source Sequence Number   | Num  |                           | Set to corresponding --SEQ of SDTM record from which ADT is selected                                                                                        |
| ASEQ          | Analysis Sequence Number | Num  |                           | Set to ADINTDT. PARAMN                                                                                                                                      |

**Table 7b ADINTDT Parameter Value Level Metadata**

| Dataset | Variable Name | Where               | Type | Derivation/Comment                                                                                                                                                                                                                        |
|---------|---------------|---------------------|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ADINTDT | ADT           | PARAMCD= 'CUTOFFDT' | Char | Assign to data base cutoff date<br>Ex: 02-05-2017                                                                                                                                                                                         |
| ADINTDT | ADT           | PARAMCD= 'DTHDT'    | Char | Select input (DD. DDDTC, is8601da. ) where DD. DDTESTCD = 'PRCDTH' and DD. DDGRPID='DEATH' before 'CUTOFFDT'                                                                                                                              |
| ADINTDT | ADT           | PARAMCD= 'LKALDT'   | Char | Select the latest date among the SDTM datasets using dates of interventions (-STDTC/-ENDTC), date of collections(-DTC), event start or end dates (-STDTC/-ENDTC )) before Cutoff date.                                                    |
| ADINTDT | ADT           | PARAMCD= 'UNCUTDT ' | Char | Select the latest date among the SDTM datasets using dates of interventions (-STDTC/-ENDTC), date of collections(-DTC), event start or end dates (-STDTC/-ENDTC )) and only populate for subjects who have the date after Date of Cutoff. |

#### **REVIEW OF ADSL (SUBJECT-LEVEL ANALYSIS DATASET):**

The variables that are in ADSL, including TRT01PN, ITTFL and STRATAR are copied to ADDTE and used in analysis. Please see the ADSL data snippet (Table 8) and metadata describing these variables of interest (Table 9). In brief, TRT01PN is a numeric version of TRT01P and TRT01P and is populated with SDTM DM.ARM values. ITTFL is set to "Y" for the subjects that are randomized. Lastly, STRATAR with values of individual stratification factor are mapped to SUPPDS domain with SUPPDS.QNAM= "STRATUM". The reviewer would be able to validate based on the description provided in metadata by checking source datasets and variables.

Table 8 ADSL Data

| USUBJID     | STRATAR | TRT01P | TRT01PN | RANDDT     | ITFL |
|-------------|---------|--------|---------|------------|------|
| XYZ-001-001 | BM+     | TPC    | 2       | 06-11-2016 | Y    |
| XYZ-001-002 | BM-     | TPC    | 2       | 02-28-2016 | Y    |
| XYZ-001-003 | BM+     | TTB    | 1       | 01-04-2016 | Y    |
| XYZ-001-004 | BM-     | TTB    | 1       | 05-25-2016 | Y    |

Table 9 ADSL metadata

| Variable Name | Variable Label                      | Type    | Codelist/Controlled Terms | Notes                                                                                                                                                                                                |
|---------------|-------------------------------------|---------|---------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| USUBJID       | Unique Subject Identifier           | Char    |                           | DM. USUBJID                                                                                                                                                                                          |
| STRATAR       | STRATA Used for Randomization       | Char    |                           | STRATAR contains the combination of values of the individual stratification factors used for randomization collected on CRF and Mapped to SUPPDS domain. SUPPDS. QVAL where SUPPDS. QNAM= "STRATUM". |
| TRT01P        | Planned Treatment for Period 01     | Char    |                           | DM. ARM                                                                                                                                                                                              |
| TRT01PN       | Planned Treatment for Period 01 (N) | Numeric |                           | Assign values as below:<br>If ADSL.TRT01P= "TTB" then<br>ADSL.TRT01PN=1;<br>else if ADSL.TRT01P= "TPC" then<br>ADSL.TRT01PN=2;                                                                       |
| RANDDT        | Date of Randomization               | Num     |                           | From DS dataset. Set RANDDT = input(DS. DSDTC, is8601da. ) where DS. DSDECOD = 'RANDOMIZATION'                                                                                                       |
| ITFL          | Intent to Treat Population Flag     | Char    |                           | Set to "Y" if ADSL. RANDDT is not missing.                                                                                                                                                           |

**REVIEW OF SDTM DOMAINS AND ANNOTATED CRFS:**

Collected data is mapped to SDTM domains. It is important to establish traceability between ADaM datasets and SDTM domains for the reviewer. As described above, the datapoint traceability between ADINTDT and SDTM domains is facilitated by using three variables: SRCDOM, SRCVAR and SRCSEQ. In ADINDT the parameter: LKALDT, Date of Last Known Alive Prior to Cutoff, is derived from multiple SDTM domains and is the latest date of contact with the subject before cutoff date. The reviewer may refer to the ADINDT program for clarity on how it is derived, and use SRCDOM, SRCVAR and SRCSEQ to trace or follow the steps all the way back to SDTM record. For example, subject XYZ-001-001 LKALDT is mapped from FADTC where FASEQ=58. Similarly, in the case of DTHDT for XYZ-001-002 it's mapped from DDDTC where DDSEQ=1. SRCDOM, SRCVAR and SRCSEQ triplet clearly establishes path from an ADaM record to SDTM record.

To further understand how the SDTM DD (Death Details) domain is populated (Table 10), the reviewer may need to refer to metadata for DD domain (Table 11) and annotated CRF (Fig 2). For illustration purposes a snapshot of DD domain along with metadata and CRF pages is provided to trace back to the collected data. The metadata below describes the variables in DD domain and refer to their origin to CRF pages. These data and metadata elements prompt the reviewer to go back to the CRF page and understand the collection and mapping methods to SDTM domains, thus providing traceability all the way to data collection.

Table 10 DD

| DOMAIN | USUBJID     | DDSEQ | DDGRPID | DDTESTCD | DDTEST                 | DDOREES                        | DDCAT                           | DDSTRESC                       | DDTDC      |
|--------|-------------|-------|---------|----------|------------------------|--------------------------------|---------------------------------|--------------------------------|------------|
| DD     | XYZ-001-002 | 1     | DEATH   | PRCDTH   | Primary Cause of Death | Malignant Neoplasm Progression | PRIMARY REPORTED CAUSE OF DEATH | Malignant neoplasm progression | 04-14-2016 |
| DD     | XYZ-001-004 | 1     | DEATH   | PRCDTH   | Primary Cause of Death | Cardiac Arrest                 | PRIMARY REPORTED CAUSE OF DEATH | Cardiac arrest                 | 08-12-2016 |

**Table 11 DD Metadata**

| Variable Name | Variable Label                       | Type    | Codelist/Controlled Terms | Origin//Method/Comment                                                           |
|---------------|--------------------------------------|---------|---------------------------|----------------------------------------------------------------------------------|
| DOMAIN        | Domain Abbreviation                  | Char    | DD                        | Two-character abbreviation for the domain.                                       |
| USUBJID       | Unique Subject Identifier            | Char    |                           | Derived. Concatenation of STUDYID and SUBJID                                     |
| DDSEQ         | Sequence Number                      | Num     |                           | Derived. Sequential number identifying records within each USUBJID in the domain |
| DDGRPID       | Group ID                             | Numeric | ddgrpid                   | CRF Pages 121                                                                    |
| DDTESTCD      | Death Detail Assessment Short Name   | Num     | ddtestcd                  | Assigned. Refer to Code list                                                     |
| DDTEST        | Death Detail Assessment Name         | Char    | ddtest                    | CRF Page 121                                                                     |
| DDCAT         | Category for Death Detail            | Char    | ddcat                     | CRF Page 121                                                                     |
| DDORRES       | Result or Finding in Original Units  | Char    |                           | CRF Page 121                                                                     |
| DDSTRESC      | Result or Finding in Standard format | Char    |                           | CRF Page 121                                                                     |
| DDTTC         | Date/Time of Collection              | Char    |                           | CRF Page 121                                                                     |

**Fig 2 Annotated CRF for DD**

**DD = Death Details**

DEATH - Form Version: 21-Sep-2015 07:38  
 Site: Subject: Subject No: **DDGRPID= 'DEATH'**

| Death                                              |                                                                    |
|----------------------------------------------------|--------------------------------------------------------------------|
| 1. Associated AE Sequence Number(s) (from AE form) | [NOT SUBMITTED] [ ] [ ] [ ] [ ] [ ] [ ]                            |
| 2. Primary Reported Cause of Death Category        | <input type="radio"/> Primary Reported Cause of Death <b>DDCAT</b> |
| 3. Primary Reported Cause of Death                 | <b>DDTEST</b> <b>DDORRES when DDTESTCD = 'PRCDTH'</b>              |
| MapDEATHTerm                                       | [NOT SUBMITTED]                                                    |
| 4. LLT                                             | [NOT SUBMITTED] [ ]                                                |
| 5. PT                                              | <b>DDSTRESC</b> [ ]                                                |
| 6. SOC                                             | [NOT SUBMITTED] [ ]                                                |
| 7. Date/Time of Death                              | <input type="checkbox"/> <b>DDTTC</b> [ ] / [ ]                    |

Traceability is also established by describing the algorithm or assigning values from the predecessor by metadata. This is called metadata traceability. Here, a well written metadata document (define.xml) is critical in order to build the metadata traceability and validate the values of analysis variables. For example, ADSL metadata (Table 9) shows the relationships of ADSL variables to its predecessor SDTM datasets and variables. As you see, ADSL.RANDDT is derived from DS (Disposition) domain with DSSTDTC value where DSDECOD= "RANDOMIZED", similarly STRATAR is populated with SUPPDS.QVAL where SUPPDS.QNAM= "STRATUM". Based on these descriptions, the reviewer can trace back to SDTM records and validate the analysis variables values. To further understand the mapping of these SDTM variables, the reviewer would review SDTM metadata and follow to data collection on the CRF form. Please see the snapshots of DS (Table 12) SUPPDS domains (Table 13) and their metadata (Table 14, Table 15) along with the annotated CRF (Fig 3). Metadata traces DSSTDTC and DSDECOD= "RANDOMIZATION" to annotated CRF. Similarly, SUPPDS.QVAL where SUPPDS.QNAM= "STRATUM" is also traced back to the CRF page.

Table 12 DS data

| DOMAIN | USUBJID     | DSSEQ | DSGRPID | DSTERM     | DSDECOD    | DSSTDTC    |
|--------|-------------|-------|---------|------------|------------|------------|
| DS     | XYZ-001-001 | 1     | SR      | RANDOMIZED | RANDOMIZED | 06-11-2016 |
| DS     | XYZ-001-002 | 1     | SR      | RANDOMIZED | RANDOMIZED | 02-28-2016 |
| DS     | XYZ-001-003 | 1     | SR      | RANDOMIZED | RANDOMIZED | 01-04-2016 |
| DS     | XYZ-001-004 | 1     | SR      | RANDOMIZED | RANDOMIZED | 05-25-2016 |

Table 13 SUPPDS data

| RDOMAIN | USUBJID     | IDVAR | IDVARVAL | QNAM    | QLABEL         | QVAL |
|---------|-------------|-------|----------|---------|----------------|------|
| DS      | XYZ-001-001 | DSSEQ | 1        | STRATUM | Stratification | BM+  |
| DS      | XYZ-001-002 | DSSEQ | 1        | STRATUM | Stratification | BM-  |
| DS      | XYZ-001-003 | DSSEQ | 1        | STRATUM | Stratification | BM+  |
| DS      | XYZ-001-004 | DSSEQ | 1        | STRATUM | Stratification | BM-  |

Table 14 DS Metadata

| Variable Name | Variable Label                          | Type    | Codelist/Controlled Terms | Origin//Method/Comment                                                           |
|---------------|-----------------------------------------|---------|---------------------------|----------------------------------------------------------------------------------|
| DOMAIN        | Domain Abbreviation                     | Char    | DS                        | Two-character abbreviation for the domain.                                       |
| USUBJID       | Unique Subject Identifier               | Char    |                           | Derived. Concatenation of STUDYID and SUBJID                                     |
| DSSEQ         | Sequence Number                         | Num     |                           | Derived. Sequential number identifying records within each USUBJID in the domain |
| DSGRPID       | Group ID                                | Numeric |                           | CRF Page 99                                                                      |
| DSTERM        | Reported Term for the Disposition Event | Num     |                           | Assigned same as DSDECOD                                                         |
| DSDECOD       | Standardized Disposition Term           | Char    | ddtest                    | CRF Page 99                                                                      |
| DSSTDTC       | Start Date/Time of Disposition Event    | Char    | ISO8601                   | CRF Page 99                                                                      |

Table 15 SUPPDS Metadata

| Variable Name | Variable Label              | Type    | Codelist/Controlled Terms | Origin//Method/Comment                                                                                                  |
|---------------|-----------------------------|---------|---------------------------|-------------------------------------------------------------------------------------------------------------------------|
| RDOMAIN       | Related Domain Abbreviation | Char    | DS                        | Two-character abbreviation for the domain of the parent record(s)                                                       |
| USUBJID       | Unique Subject Identifier   | Char    |                           | Derived. Concatenation of STUDYID and SUBJID                                                                            |
| IDVAR         | Identifying Variable        | Num     |                           | Assigned. Identifying variable in the parent dataset that identifies supplemental record(s). DSSEQ in the present case. |
| IDVARVAL      | Identifying Variable Value  | Numeric |                           | Assigned. The value of DSSEQ                                                                                            |
| QNAM          | Qualifier Variable Name     | Num     |                           | Assigned as "STRATUM"                                                                                                   |
| QLABEL        | Qualifier Variable Label    | Char    |                           | Assigned as "Stratification"                                                                                            |
| QVAL          | Data Value                  | Char    |                           | Value of QNAM                                                                                                           |

Fig 3 Annotated CRF for DS/SUPPDS

DS = Disposition

SUBJECT RANDOMIZATION - Form Version: 21-Sep-2015 07:47  
 Site: Subject: Subject No: DSGRPID = 'SR'

---

**SUBJECT RANDOMIZATION**

|    |                       |                                      |  |                                                        |         |
|----|-----------------------|--------------------------------------|--|--------------------------------------------------------|---------|
| 1. | Date of Randomization | DSDECOD='RANDOMIZATION'              |  |                                                        | DSSTDTC |
| 2. | Stratification        | QVAL when QNAM = 'STRATUM' in SUPPDS |  | <input type="radio"/> BM+<br><input type="radio"/> BM- |         |

**CONCLUSION:**

Here we present OS from oncology trials to provide a practical example using a combination of CDISC standards to establish traceability for regulatory review. The step-by-step examples presented in the paper begin with the CSR OS Table or “the end” product and follow through to the beginning using the CDISC standards tools of ARM, Analysis data, Analysis Data Metadata, SDTM data, SDTM Metadata and annotated CRFs. In preparing a submission, it is important to create clear ARM and metadata for datasets (SDTM, ADaM) to help the reviewer effectively trace back from the analysis results presented in the CSR tables to CRF data. In summary, traceability is the key and should be established throughout data life cycle, thus helping reviewers to effectively review and validate the regulatory submission.

**REFERENCES:**

Please refer to <https://www.cdisc.org/> for all CDISC documents discussed in this paper.

**ACKNOWLEDGMENTS:**

The authors would like to thank Merck’s Oncology Programming management team, and Merck Global Clinical Data Standards for their support and encouragement. The authors also greatly appreciate the Merck oncology standard team for standardizing data structures for ADTTE and ADINTDT.

**CONTACT INFORMATION:**

Your comments and questions are valued and encouraged. You may contact the authors at:

Madhusudhan Reddy Papasani  
Merck & Co, Inc.  
madhusudhan.papasani@merck.com

Swathi Kotla  
Merck & Co, Inc.  
swathi.kotla@merck.com

Elisabeth Pyle  
Merck & Co, Inc.  
lisa\_pyle@merck.com