

Let Us Trace Back to the Beginning from the End

Madhusudhan Reddy Papasani, Merck & Co., Inc., Upper Gwynedd, PA

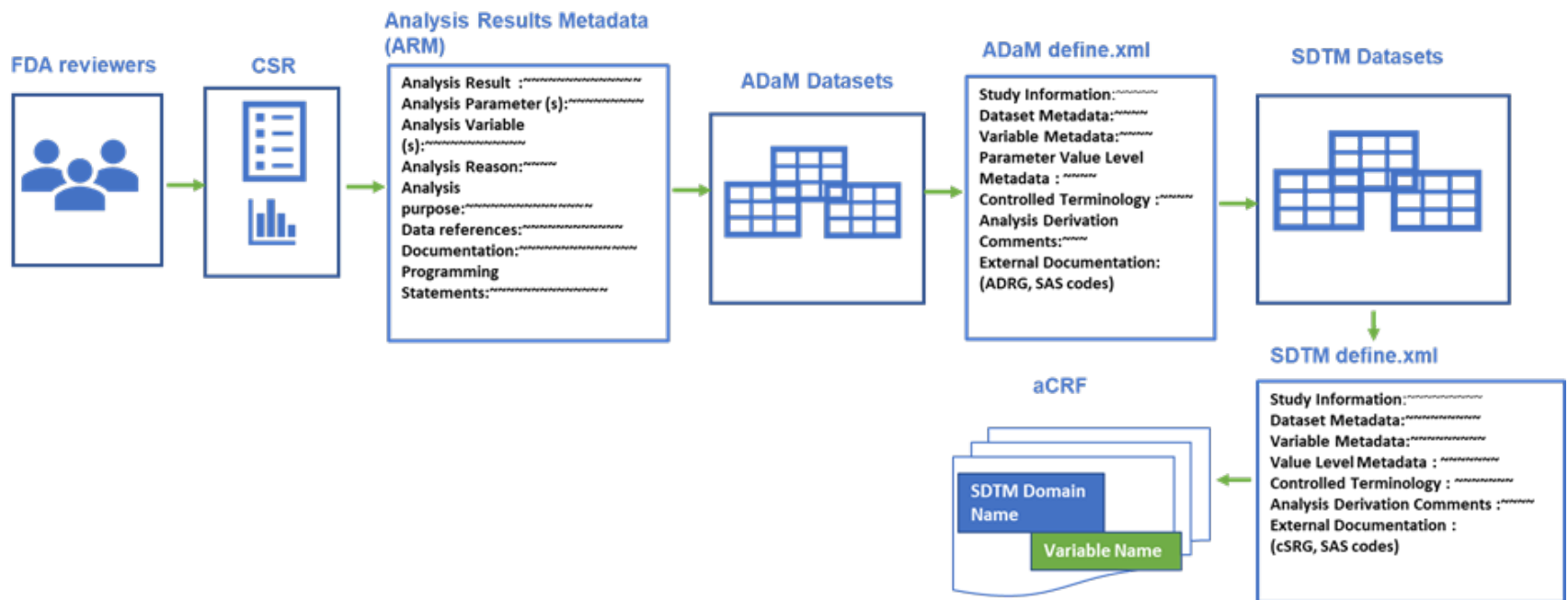
Swathi Kotla, Merck & Co., Inc., Upper Gwynedd, PA

Elisabeth Pyle, Merck & Co., Inc., Upper Gwynedd, PA

Introduction and Purpose

- 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
- 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
- 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 are presented, traced and highlighted

The relationships of CDISC standards in the FDA review process starting from CSR



Overall Survival (OS)

- In oncology trials, overall survival (OS) is a key efficacy endpoint.
- 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.
- 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. Estimation of the hazard ratio will be performed using a stratified Cox regression model. The treatment difference in survival will be assessed by the stratified log-rank test.

An Example CSR table summarizing OS

Analysis of Overall Survival
All Subjects (ITT Population)

	TTB (N = ###)	TPC (N = ###)
Kaplan-Meier Estimates (Months)[†]	##. # (##. #, ##. #)	##. # (##. #, ##. #)
Median (95% CI)		
vs TPC		
Hazard Ratio (95% CI) [‡]	#. ## (#. ##, #. ##)	
p-value [§]	0. #####	

[†] From product-limit (Kaplan-Meier) method for censored data.

[‡] 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-).

[§] p-value based on stratified log-rank test.

Quick Validation of Results Using ARM

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-values: <pre>proc lifetest data = ADTTE; where TRT01PN in (1,2); time AVAL*CNSR(1); strata STRATAR/group=TRT01PN; ods output HomTests=pval(where=(test='Log-Rank')); run;</pre> Further reference: by checking program: SAS Version: [SAS Version 9. 4] SAS program: e0or0anal0tte0ver0osgen.txt

Review of ADTTE dataset and Metadata

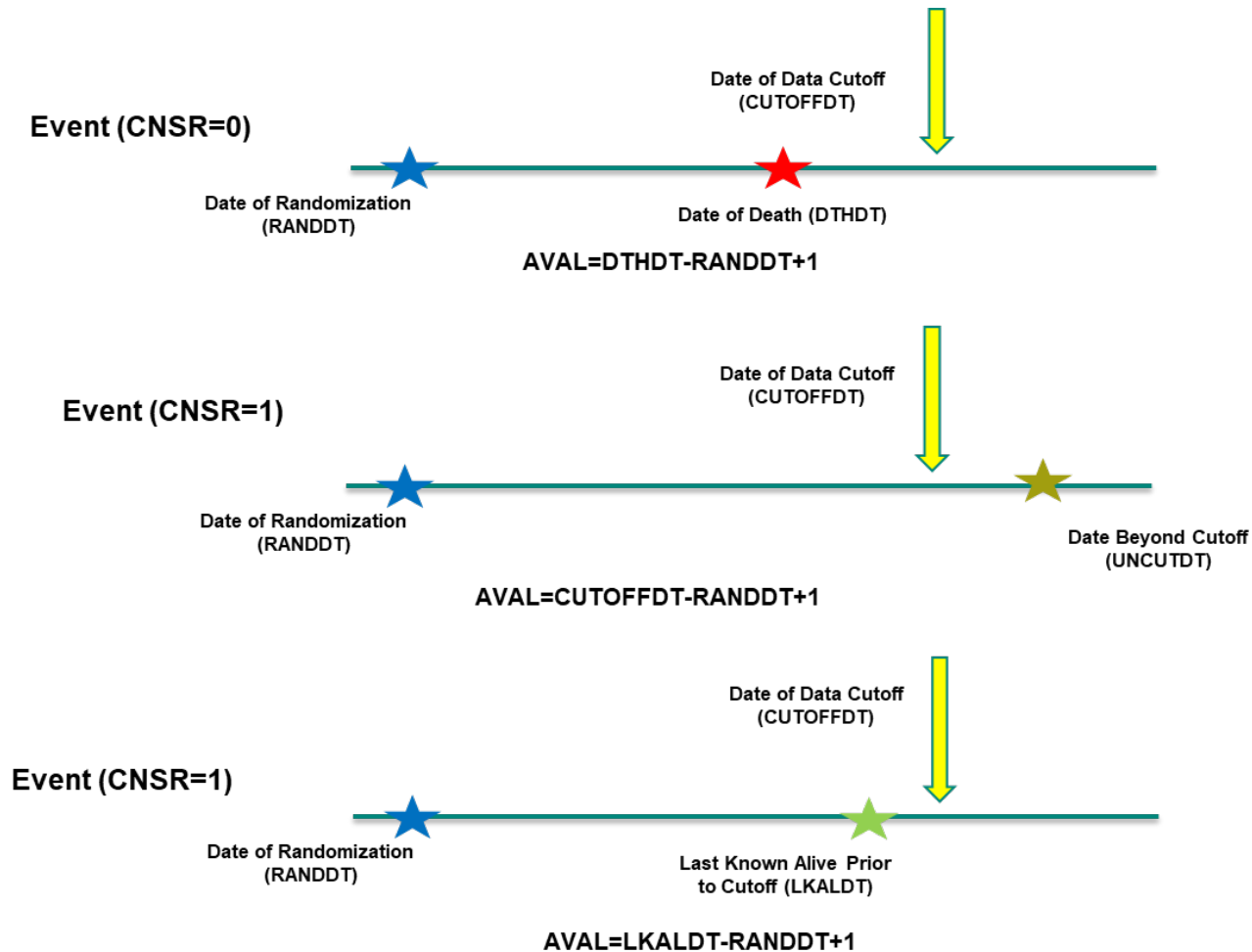
ADTTE DATA

USUBJID	PARAMCD	STARTDT	ADT	AVAL	SRCDOM	SRCVAR	SRCSEQ	CNSR	STRATAR	ITTFL	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

ADTTE 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		Analysis date for the event (in this case Death) or for Censoring. If a subject died on or prior to the data cutoff date, then the analysis date is the date of death (Event). Else if a subject is either alive or dead after cutoff date then the subject is censored on the cutoff date (Censor); Please note that uncut date is only populated for the subjects who have last known dates beyond cutoff date. 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).
AVAL	Analysis Value	Num		Time from randomization date to death date or censoring date. Programming logic: ADTTE. AVAL=ADTTE. ADT-ADTTE. STARTDT+1
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

Concepts behind deriving AVAL and CNSR in ADTTE



ADTTE Program –snapshot

Transpose ADINDT
 Transposed ADINDT as a temporary SAS
 Derive/Assign PARAM, ADT, AVAL, SRCDOM, SRCVAR

```

•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;
  
```

```

**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;
  
```

```

**Merging dates and
sequence datasets;
data __adintdt2;
merge dates seq;
by usubjid;
run;
  
```

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	

```

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_lkal2t;
end;
aval=adt-startdt+1;
format startdt yymmdd10. ;
run;
  
```

Review of ADINTDT dataset and Metadata

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

ADINDT 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"
PARAM	Parameter	Char		Set ADINTDT. PARAM to the following Values: Date of Data Cutoff, Date of Death, Date of Last Known Alive Before Data Cutoff, Date Beyond Cutoff
PARAMN	Parameter (N)	Num		Assign Values as below for ADINTDT.PARAMCDs: 1 for CUTOFFDT, 2 for DTHDT, 3 for LKALDT, 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

Review of ADINTDT dataset and Metadata-Continued

ADINTDT Parameter Value Level Metadata

Dataset	Variable Name	Where	Type	Derivation/Comment
ADINTDT	ADT	PARAMCD= 'CUTOFFDT'	Char	Assign to data base cutoff date 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.

ADINTDT Program –snapshot

```

*Date of Last Known Alive Prior to Cutoff;
%let cutoff_dt= 02-05-2017;
data dts;
  set ex (in=xs keep=usubjid domain exstdtc exseq rename=(exstdtc=_dtc exseq=srcseq))
    ex(in=xe 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 trdrtc trseq rename=(trdrtc=_dtc trseq=srcseq))
    rs (in=rs keep=usubjid domain rsdrtc rsseq rename=(rsdrtc=_dtc rsseq=srcseq))
    lb (in=lb keep=usubjid domain lbdrtc lbseq rename=(lbdrtc=_dtc lbseq=srcseq))
    vs(in=vs keep=usubjid domain vsdrtc vsseq rename=(vsdrtc=_dtc vsseq=srcseq))
    fa(in=fa keep=usubjid domain fadrtc faseq fagrpid faorres rename=(fadrtc=_dtc
      faseq=srcseq) where=(fagrpid='SFU' and upcase(faorres)='ALIVE'))
    ds(in=ds keep=usubjid domain dsdrtc dsseq dsgrpid rename=(dsdrtc=_dtc
      dsseq=srcseq) where=(dsgrpid='SR'))
    ss(in=ss keep=usubjid domain ssdrtc ssseq rename=(ssdrtc=_dtc ssseq=srcseq));
  if _dtc>' ' and input(_dtc, ?? yymmdd10. ) and adt<="&cutoff_dt"d;
  length srcdom srcvar $8;
  srcdom=domain;
  if xs then srcvar='EXSTDTC';
  else if xe then srcvar='EXENDTC';
  else if sv then srcvar='SVSTDTC';
  else if tr then srcvar='TRDRTC';
  else if rs then srcvar='RSDRTC';
  else if lb then srcvar='LBDRTC';
  else if vs then srcvar='VSDRTC';
  else if fa then srcvar='FADRTC';
  else if ds then srcvar='DSDRTC';
  else if ss then srcvar='SSDRTC';
  adt=input(_dtc, ?? yymmdd10. );
  format adt yymmdd10. ;run;

*Sort the resulting dataset and select the latest date
proc sql;
create table __dts_tmp as
select d. * from dts as d
order by usubjid, adt, srcdom desc, srcvar desc, srcseq;
quit;
data lnaldt ;
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;

```

```

*Date of Data Cutoff;
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 ddtested='PRCDTH' and
    dddtc ne ' ';
  adt=input (dddtc,yymmdd10. );
  srcdom= 'DD'
  srcvar= 'DDDTC';
  srcseq= DDSEQ;
  param='Date of Death';
  paramcd='DTHDT';
  paramn=2;
  if adt <="&cutoff_dt"d;
run;

```

```

*Create ADINTDT;
data adintdt;
set cutoffdt dthdt lnaldt uncutdt;
  aseq=paramn;
run;

```

Review of ADSL dataset and Metadata

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: If ADSL.TRT01P= "TTB" then ADSL.TRT01PN=1; else if ADSL.TRT01P= "TPC" then ADSL.TRT01PN=2;
RANDDT	Analysis Date	Num		From DS dataset. Set RANDDT = input(DS. DSDTC, is8601da.) where DS. DSDECOD = 'RANDOMIZATION'
ITTFLL	Intent to Treat Population Flag	Char		Set to "Y" if ADSL. RANDDT is not missing.

ADSL Data

USUBJID	STRATAR	TRT01P	TRT01PN	RANDDT	ITTFLL
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

Review of SDTM datasets and CRFs

DD Data

DOMAIN	USUBJID	DDSEQ	DDGRPID	DDTESTCD	DDTEST	DDOREES	DDCAT	DDSTRESC	DDDTCT
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

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
DDDTCT	Date/Time of Collection	Char		CRF Page 121

aCRF

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	☐ Primary Reported Cause of Death DDCAT
3. Primary Reported Cause of Death	DDTEST DDORRES when DDTESTCD = 'PRCDTH'
MapDEATHTerm	[NOT SUBMITTED]
4. LLT	[NOT SUBMITTED]
5. PT	DDSTRESC
6. SOC	[NOT SUBMITTED]
7. Date/Time of Death	DDDTCT

Review of SDTM datasets and CRFs Continued

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

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-

DS Metadata

Variable Name	Variable Label	Type	Codelist/Controlled Terms	Origin//Method/Comment
DOMAIN	Domain Abbreviation	Char	DS	Two-character abbreviation for

SUPPDS Metadata

Variable Name	Variable Label	Type	Codelist/Controlled Terms	Origin//Method/Comment
---------------	----------------	------	---------------------------	------------------------

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 ☐ BM+ ☐ BM-
QVAL when QNAM = 'STRATUM' in SUPPDS

DSTERM	Reported term Num 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

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

Two-character abbreviation for the main of the parent record(s)
 derived. Concatenation STUDYID and BJID
 assigned. Identifying variable in the parent dataset that identifies supplemental

record(s). DSSEQ in the present case.

Conclusion

- We presented 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
- 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.

Thank You!

