FDA U.S. FOOD & DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION & RESEARCH

CDISC JAPAN INTERCHANGE 2021

CDER'S EXPERIENCE WITH THE ADAM TRACEABILITY ASSESSMENT AND COMMON DATA QUALITY ISSUES

Jesse Anderson

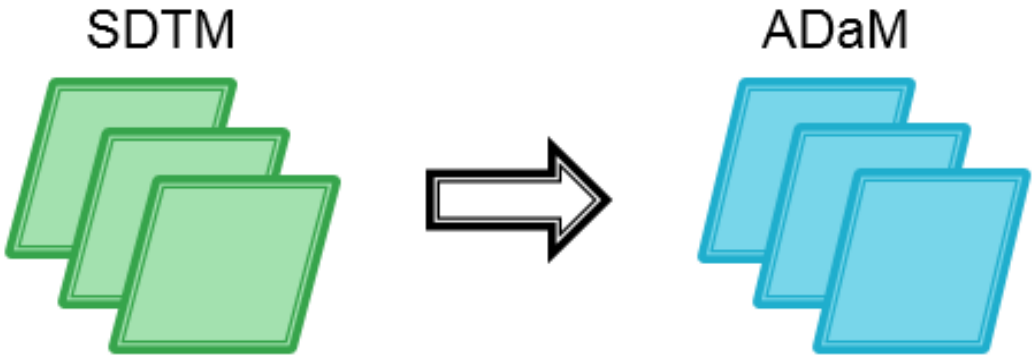
Lead, Data Standards and JumpStart Service
Division of Regulatory Review and Research
Office of Computational Science, CDER, FDA

FDA Disclaimer

The views and opinions presented here represent those of the speaker and should not be considered to represent advice or guidance on behalf of the U.S. Food and Drug Administration.

SDTM to ADaM Traceability Overview

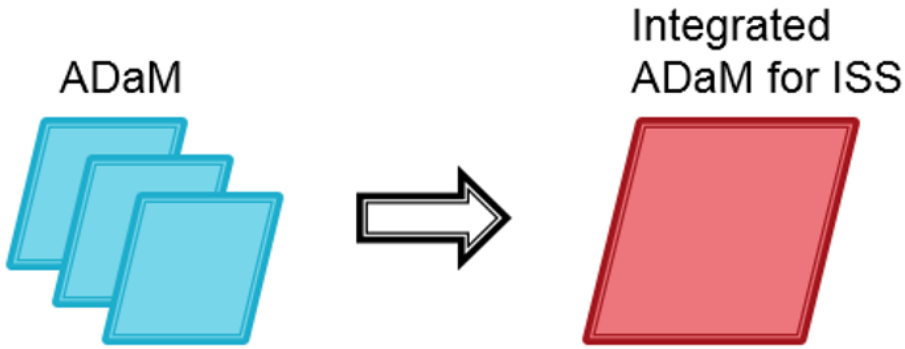
- The Study Data Tabulation Model (SDTM) to the Analysis Data Model (ADaM) Traceability Assessment traces SDTM data to the ADaM data to identify differences that may exist between key datasets and to assess their level of impact



- Datasets compared include four topic areas:
  - Demographics (SDTM.DM and ADaM.ADSL)
  - Adverse Events (SDTM.AE and ADaM.ADAE)
  - Laboratory (SDTM.LB and ADaM.ADLB)
  - Vital Signs (SDTM.VS and ADaM.ADVS)

Individual Study ADaM to ISS ADaM Traceability Overview

- The traceability assessment was expanded to reconcile record count and compare key variables between the individual study ADaM datasets and the Integrated Summary of Safety (ISS) ADaM datasets for the same topic areas: Demographics, Adverse Events, Laboratory, and Vital Signs
  - In this assessment, the same key variable concepts listed in subsequent slides by each topic area are checked across topic areas. Since this assessment compares ADaM datasets, the SAS program compares analysis variables (e.g., xxADTM vs. xxADTM)



- This assessment is useful for reviewers who plan to use the ISS data in their analysis

Two Key Areas of Assessment

- This assessment, targeted toward reviewers who plan to use ADaM data in their analysis, focuses on two key areas:
  - **Record Count Reconciliation:** Summarizing which records that exist in one dataset but not the other and understanding the reason behind these differences
  - **Variable Comparison:** For records that exist in both datasets, comparing key variables to identify differences in values and assessing the impact of these differences

- SAS programs are used to perform the record count reconciliation and subsequent variable comparisons
  - The SAS programs are customizable to adjust for variations in variable names and formats

Record Count Reconciliation

- It is common to see records submitted in one dataset but not the other
  - For example, derived records may exist in ADaM datasets that are not provided in the tabulated SDTM datasets. Similarly, events or findings for Screen Failure or Not Assigned subjects may exist in the SDTM datasets that are not preserved in the ADaM datasets
  - Record counts may match 1:1, distinct records may exist in both datasets (as shown in Figure 1), or distinct records may only exist in one dataset (as shown in Figure 2)

Figure 1

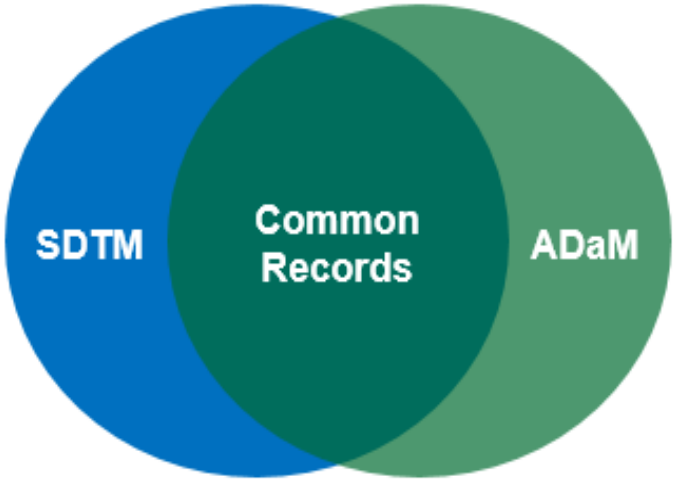
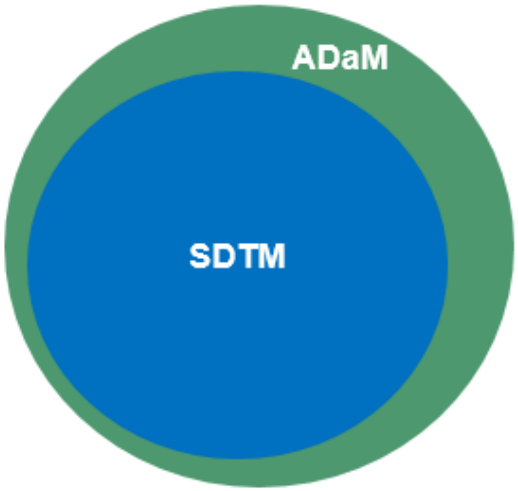


Figure 2



- **The objective of the reconciliation is to categorize records that are not submitted in both datasets and assess the impact of these discrepancies**

Variable Comparison

- After understanding which records are submitted in both datasets, these records are mapped using Unique Subject Identifier (USUBJID) for Demographics, and both USUBJID and Sequence Number (xxSEQ) for Adverse Events, Laboratory, and Vital Signs
- The process of mapping records to compare values highlights the importance of submitting the **Sequence Number** variable in findings and events datasets for both SDTM and ADaM
  - Missing or re-sequenced sequence number values have resulted in Information Requests (IRs) being submitted to applicants in past submissions
- Subsequent slides summarize variables checked and commonly observed findings across each topic area

Adverse Events Record Count Differences

- Examples of observed findings related to record count differences between the SDTM Adverse Events (AE) domain and the ADaM Adverse Events Analysis Dataset (ADAE) (or ADAE and ISS ADAE) and their level of impact are listed in the table below:

Key
No Impact
Low Impact
High Impact

Finding	Impact
Events that occurred after the ADAE-defined TEAE cutoff date captured in the AE domain are not preserved in the ADAE dataset	**Low impact**: The treatment emergent flag in the SUPPAE domain or in the ADAE dataset can be leveraged to assess treatment emergent adverse events (TEAEs)
Events that occurred prior to the screening period are captured in the AE domain and are not preserved in the ADAE dataset	**Low impact**: It is expected that events captured prior to the screening period are captured in the Medical History (MH) domain
Unexplained removal of adverse events in the SDTM or ADaM datasets without a clear explanation provided in the documentation	**High impact**: The unexplained removal of events denoting treated subjects call into question the data practices of the applicant

Adverse Events: Variables Checked

- Below is a table containing the variables checked between the SDTM AE domain and the ADaM ADAE dataset. For ISS Traceability, the ADAE variable is checked between the individual study and the ISS dataset:

Category	AE Variable Name	AE Variable Label	ADAE Variable Name	ADAE Variable Label
MedDRA Coding	AEDECOD	Dictionary-Derived Term	AEDECOD	Dictionary-Derived Term
	AEBODSYS	Body System or Organ Class	AEBODSYS	Body System or Organ Class
Severity/ Toxicity	AESEV	Severity/Intensity	ASEV	Analysis Severity/Intensity
	AETOXGR	Standard Toxicity Grade	ATOXGR	Analysis Standard Toxicity Grade
Seriousness and Seriousness Qualifiers	AESER	Serious Event	ASER	Analysis Serious Event
	AESHOSP	Requires or Prolongs Hospitalization	AESHOSP	Requires or Prolongs Hospitalization
	AESDTH	Results in Death	AESDTH	Results in Death
	AESLIFE	Is Life Threatening	AESLIFE	Is Life Threatening
	AESMIE	Other Medically Important Serious Event	AESMIE	Other Medically Important Serious Event
	AESCONG	Congenital Anomaly or Birth Defect	AESCONG	Congenital Anomaly or Birth Defect
	AESDISAB	Persistent or Significant Disability/Incapacity	AESDISAB	Persistent or Significant Disability/Incapacity
	AESCAN	Involved Cancer	AESCAN	Involved Cancer
	AESOD	Occurred with Overdose	AESOD	Occurred with Overdose

Adverse Events: Variables Checked

- Below is a table continuing the list of variables checked between the SDTM AE domain and the ADaM ADAE dataset. For ISS Traceability, the ADAE variable is checked between the individual study and the ISS dataset:

Category	AE Variable Name	AE Variable Label	ADAE Variable Name	ADAE Variable Label
Date/ Time	AESTDTC	Start Date/Time of Adverse Event	ASTDT/ ASTDTM	Analysis Start Date/ Analysis Start Date & Time
	AEENDTC	End Date/Time of Adverse Event	AENDT/ AENDTM	Analysis End Date/ Analysis End Date & Time
Other	AEACN	Action Taken with Study Treatment	AEACN	Action Taken with Study Treatment
	AEOUT	Outcome of Adverse Event	AOUT	Analysis Outcome of Adverse Event
	AEREL	Causality	AREL	Analysis Causality

- A treatment emergent flag check is also completed if a treatment emergent flag variable is submitted in the Supplemental Adverse Events (SUPPAE) domain:

Category	SUPPAE Variable Name	SUPPAE Variable Label	ADAE Variable Name	ADAE Variable Label
TEAE Flag	AETRTEM	Treatment Emergent Flag	TRTEMFL	Treatment Emergent Flag

Adverse Events: Variable Differences

- Examples of observed findings related to value differences across key variables between the SDTM AE domain and the ADaM ADAE dataset (or ADAE and ISS ADAE) and their level of impact are listed in the table below:

Key
No Impact
Low Impact
High Impact

Finding	Impact
Date and/or time values differ between datasets	**Low impact:** Missing or partial dates in SDTM.AE are imputed in ADaM.ADAE as defined in the documentation such as the ADRG or the SAP. Partial dates in the SDTM.AE domain may be missing either day, day and month, or day, month, and year (i.e., missing), which is imputed as a full date in the ADaM ADAE dataset. Mismatching dates may impact duration analysis and assessing exposure and relation to study drug
	High impact: Mismatching nonpartial date values that do not align with the source identified in the Define.xml file or other documentation. Mismatching dates may impact duration analysis and assessing exposure and relation to study drug
Difference in Treatment Emergent Flag	**Low impact:** Discrepancies are due to differences in the flag derivation between SDTM and ADaM (or ADaM and ISS ADaM) as defined in the SDTM and ADaM Define.xml file, SDRG and ADRG, or SAP documentation
	High impact: Mismatching records that do not align with the derivation provided in the documentation call into question the data practices of the applicant

Difference in Treatment Emergent Flags

- Below is an illustrative example of differences in treatment emergent flag due to differences in flag derivation between SDTM and ADaM:

DM			AE			SUPPAE	ADAE		
USUBJID	RFXSTDTC	RFXENDTC	AEDECOD	AESTDTC	AEENDTC	AETRTEM	ASTDTM	AENDTM	TRTEMFL
Subject A	2020-01-02 T12:00:00	2020-01-02 T12:00:00	Fatigue	2020-01-02 T10:00:00	2020-01-08 T:10:00:00	Y	2020-01-02 T10:00:00	2020-01-08 T:10:00:00	
Subject B	2020-01-02 T12:00:00	2020-03-02 T12:00:00	Nausea	2020-01	2020-02-05 T:10:00:00	N	2020-01-02 T12:00:00	2020-02-05 T:10:00:00	Y
Subject C	2020-01-02 T12:00:00	2020-03-02 T12:00:00	Dizziness	2020-06-12 T12:00:00	2020-08-15 T08:30:00	Y	2020-06-12 T12:00:00	2020-08-15 T08:30:00	

- **Subject A**: The derivation of SUPPAE.AETRTEM only takes the date component of AESTDTC into account, ignoring the time component. Since the derivation of ADAE.TRTEMFL takes both date and time into account, the treatment emergent flag is null in ADAE
- **Subject B**: A partial date in AESTDTC results in AETRTEM set equal to 'N'. However, after imputing start and end date in the ADAE dataset, the adverse event now falls within the treatment emergent window in ADAE and is set equal to 'Y'
- **Subject C**: The cutoff date within the treatment emergent flag derivations differ between SUPPAE.AETRTEM and ADAE.TRTEMFL. Here, assume the SDTM Define.xml lists a derivation for SUPPAE.AETRTEMFL as '*on or after start of first exposure to treatment*', whereas the ADaM Define.xml lists a derivation for ADAE.TRTEMFL as '*on or after start of first exposure to treatment, up to 30 days past the last exposure to treatment*'

Laboratory/Vital Signs Record Count Differences

- Examples of observed findings related to record count differences between the SDTM Laboratory/Vital Signs (LB/VS) domain and the ADaM Laboratory Results/Vital Signs Analysis (ADLB/ADVS) dataset (or ADLB/ADVS and ISS ADLB/ADVS) and their level of impact are listed in the table below:

Key
No Impact
Low Impact
High Impact

Finding	Impact
Certain tests submitted in one dataset but excluded from the other	**Low impact:** Tests that are not needed to perform relevant analysis may be excluded from analysis datasets
Derived records in ADaM are not submitted in the SDTM datasets, or derived records in individual study ADaM differ from derived records submitted in ISS ADaM	**Low impact**: It is expected for derived records to only exist in the ADaM dataset. Records may be derived in ISS datasets which yield differences across derived records (e.g., if the individual study ADaM dataset derives LVPD (i.e. baseline) records but the ISS ADaM dataset derives End of Study (EOS) records
Unexplained removal findings in the SDTM or ADaM datasets without a clear explanation provided in the documentation	**High impact**: The unexplained removal of findings pertaining to treated subjects call into question the data practices of the applicant

Laboratory/Vital Signs: Variables Checked

- Below is a table containing the variables checked between the SDTM LB/VS domain and the ADaM ADLB/ADVS dataset. For ISS Traceability, the ADLB/ADVS variable is checked between the individual study and the ISS dataset:

LB/VS Variable Name	LB/VS Variable Label	ADLB/ADVS Variable Name	ADLB/ADVS Variable Label
xxSTRESN	Numerical Result/Finding in Standard Units	AVAL	Analysis Value
xxSTRESC	Character Result/Finding in Standard Format	AVALC	Analysis Value (Character)
xxSTRESU	Standard Units	AVALU	Analysis Units
xxTEST	Lab/Vital Signs Test or Examination Name	PARAM	Parameter
xxCAT	Category for Lab/Vital Signs Test	$PARCAT_y$	Parameter Category y
VISIT	Visit Name	AVISIT	Analysis Visit
xxDTC	Date/Time of Specimen Collection/ Measurements	ADT/ADM	Analysis Date/ Analysis Time
xxBLFL	Baseline Flag	ABLFL	Baseline Record Flag

Laboratory/Vital Signs: Variable Differences

- Examples of observed findings related to value differences across key variables between the SDTM LB/VS domain and the ADaM ADLB/ADVS dataset (or ADLB/ADVS and ISS ADLB/ADVS) and their level of impact are listed in the table below:

Key
No Impact
Low Impact
High Impact

Finding	Impact
Difference in numerical value due to value imputation, unit conversion, or rounding	**Low impact:** Low impact when differences in values are documented but results will be different depending on the dataset used. A common example of value imputation includes null SDTM xxSTRESN imputed in ADaM AVAL as the numerical component of xxSTRESC, which may include inequality symbols (e.g., '<3'). If value imputation or other reasons for differences are unclear, this may have higher impact on the review
Difference in visit value	**Low impact:** Differences are often due to derivation differences in SDTM vs. ADaM or ADaM vs. ISS ADaM, which may be sourced from different variables such as study day, baseline flags, or time windows. Note that differences in visit are common per the ADaMIG v1.1: *'AVISIT is a derived field and does not have to map to VISIT from the SDTM.'* Understanding how different visit variables are derived aid the reviewer in deciding which variable to use in analysis if using visit is desired

Laboratory Value Differences

- Below is an illustrative example of differences in numerical result value due to value imputation:

LB domain				ADLB dataset		
USUBJID	LBTEST	LBSTRESC	LBSTRESN	AVALC	AVAL	AVALU
Subject A	Bilirubin	<0.2		<0.2	0.2	mg/dL
Subject B	Bilirubin	<0.3		<0.3	0.15	mg/dL

  - **Subject A**: Missing LBSTRESN is imputed as the numerical component of LBSTRESC
  - **Subject B**: Missing LBSTRESN is imputed as the numerical component of LBSTRESC divided in half
- The methodology behind value imputation provided by the applicant is not always clear

Laboratory/Vital Signs: Variable Differences

- Examples of observed findings related to value differences across key variables between the SDTM LB/VS domain and the ADaM ADLB/ADVS dataset (or ADLB/ADVS and ISS ADLB/ADVS) and their level of impact are listed in the table below:

Key
No Impact
Low Impact
High Impact

Finding	Impact
Difference in Baseline Flag	**Low impact:** Discrepancies are due to differences in the flag derivation between SDTM and ADaM (or ADaM and ISS ADaM) as defined in the SDTM and ADaM Define.xml file, SDRG and ADRG, or SAP documentation
	Low impact: Baseline flag values being dropped in ADaM for select tests, such as a pregnancy test records in the LB domain
	Low impact: Multiple baseline flags assigned in SDTM may be removed in ADaM, so that each subject and test only have a maximum of one baseline record. Conversely, a subject and test may have multiple baseline flags in ADaM due to assigning multiple flags for both local and central laboratory results or due to multiple BASETYPE values to account for safety and efficacy analysis needs
	Low impact: Baseline flags that were not derived in the SDTM dataset but are derived in the ADaM dataset
	High impact: Mismatching records that do not align with the derivation provided in the documentation call into question the data practices of the applicant

Laboratory Baseline Record Flag Differences

- Below is an illustrative example of differences in baseline record flag due to derivation differences:

DM domain		LB domain						ADLB dataset	
USUBJID	RFXSTDTC	LBTEST	LBSTRESN	LBSTRESU	LBDTC	LBDY	LBBLFL	AVAL	ABLFL
Subject B	2020-01-08 **T08:30:00**	Bilirubin	0.15	mg/dL	2020-01-01 T12:00:00	-7	Y	0.15	
		Bilirubin	0.18	mg/dL	2020-01-08 **T08:00:00**	1		0.18	Y

The baseline record flag found in the LB domain is assigned to the screening record that took place a week before treatment began

The baseline record flag found in the ADLB dataset aligns with the record that occurred on the same day but at a time prior to the beginning of treatment

- In this example, the derivation in ADaM takes time component into account whereas SDTM does not
- Differing numerical value for baseline records may result in shift tables and other baseline vs. postbaseline analysis will return different results if using SDTM rather than ADaM datasets
- The methodology behind value imputation provided by the applicant is not always clear

Conclusion

- Traceability is a key area of analysis – an understanding of differences between SDTM, ADaM, and ISS ADaM allows reviewers to confidently use these submitted data packages throughout the review of their application
- The assessment aims to help the reviewer understand the impact of differences observed in the datasets for their application
  - Findings of higher impact, where the differences in datasets are unclear and may call into question the data practices of the applicant, also may result in an Information Request (IR) being filed
- The JumpStart team reports this wide range of observed traceability findings to inform applicants of the need for thorough data and documentation. A comprehensive submission increases the efficiency and timeliness of the review. Two essential components include:
  - Thorough documentation including derivations and other key information submitted in the Reviewer's Guide, SAP, and/or Define.xml File
  - Sequence Numbers submitted in both SDTM and ADaM datasets that map records between comparable findings and events datasets (e.g., AE and ADAE)

Thank You!

Jesse Anderson

Lead, Data Standards and Nonclinical Program

Division of Regulatory Review and Research

Office of Computational Science

Office of Translational Sciences

FDA CDER


FDA
U.S. FOOD & DRUG
ADMINISTRATION