Overseas Regulatory CDISC Submission Experience

Presented by Tina Pyo
Head of Clinical Statistics Department, CELLTRION

Meet the Speaker

Tina Pyo

Title: Head of Clinical Statistics Department

Organization: CELLTRION

Tina Pyo has been working in the industry since January 2011, giving her over 13 years of experience. She spent two and a half years as a biostatistician at DreamCIS (CRO), and since then, she has progressed through roles at CELLTRION, serving as a Biostatistician and Team Leader, and is now the Head of Clinical Statistics Department.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

- *The speak has no real or apparent conflicts of interest to report.*

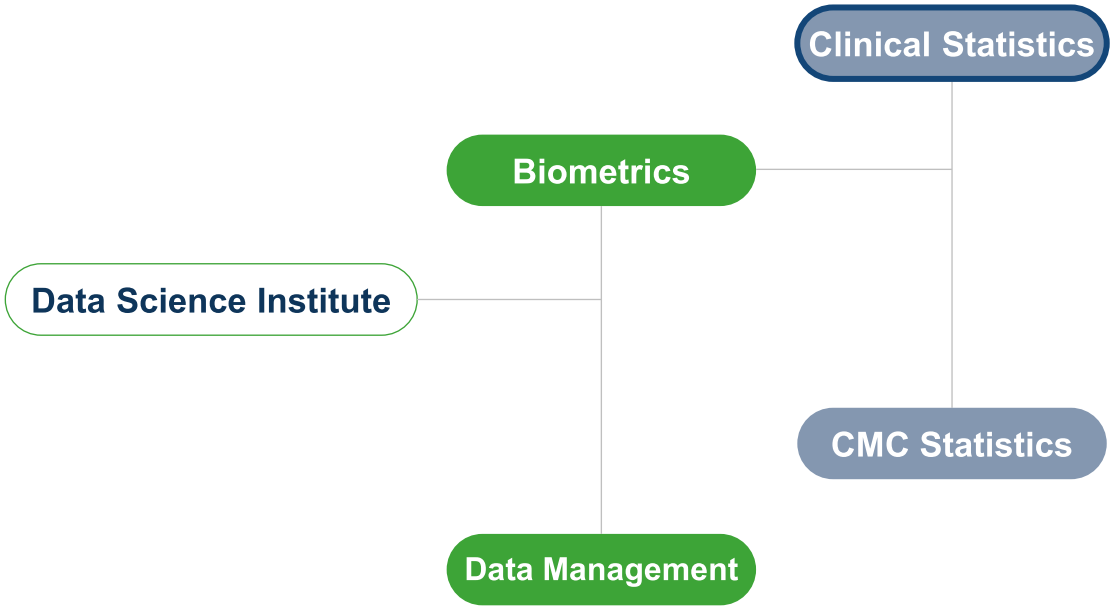
Agenda

1. Introduction of Celltrion Biometrics
2. Important Consideration for ADaM
3. CDISC submission case studies
4. Summary and Suggestions

Introduction of Celltrion Biometrics

Biometrics in Celltrion

Biometrics in Celltrion



13 years of Clinical Stat Team:

- **CDISC-compliant submission: 20+ studies submission experience in 10+ products**
- **Most of works in-house**

Biometrics in Celltrion

Primary Goal of CDISC package preparation

Internal

Meet the submission timeline

Regulatory

Quality and integrity

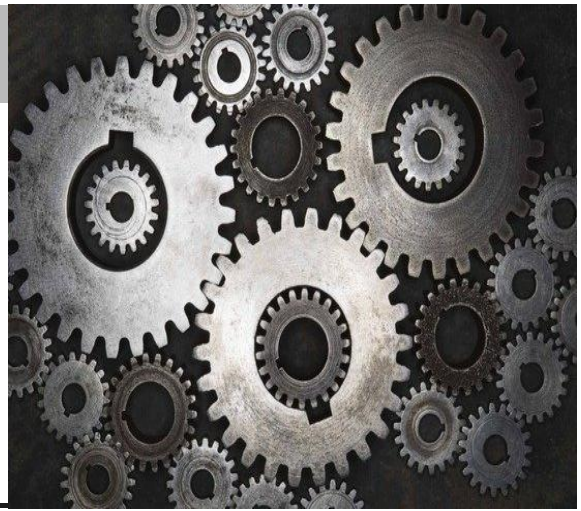
"The applicant submitted data files of acceptable quality and it was possible to reproduce the primary analysis dataset."

"There are no concerns regarding data quality and integrity."

Biometrics in Celltrion

Meticulous and Planning in advance

- ✓ Technical skillset
- ✓ Necessary tools in place
- ✓ Motivation for the work
- ✓ Collaborative communication in early stage
- ✓ Understanding of regulatory perspective



Important Consideration for ADaM

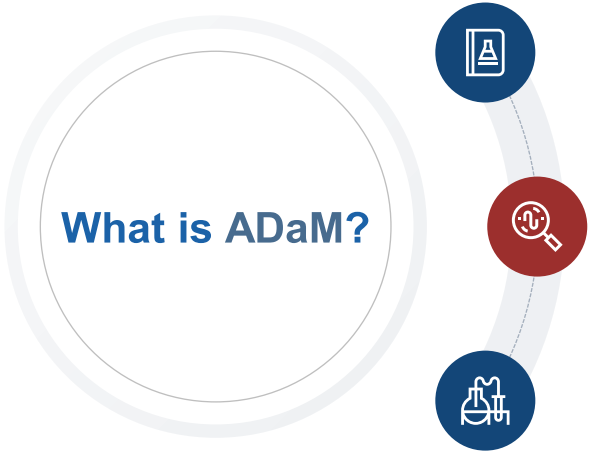
Introduction to ADaM

Fundamental Principle

Misunderstandings about ADaM Implementation

Technical Considerations

Introduction to ADaM: Definition and Goal



Analysis Data Model (ADaM) defines dataset and metadata standards that support:

- efficient generation, replication, and review of clinical trial statistical analyses
- traceability among analysis results, analysis data, and data represented in the SDTM.

ADaM is not just Analysis Dataset. (Metadata is a key component of ADaM)

ADaM is one of the required standards for data submission to FDA and PMDA.

Introduction to ADaM: Definition and Goal

Provides clear, unambiguous communication of the statistical aspects of the trial through the structure and content of the analysis datasets.

Allows recipients of analysis datasets to understand data lineage from collection to analysis to results

The ADaM standard has been developed to meet the needs of the industry and regulatory agencies.
Supports efficient generation, replication, and review of analysis results.

Introduction to ADaM: Documents

- **Documents from CDISC**

Analysis Data Model v2.1	Basic principles to apply to all analysis datasets.
Analysis Data Model Implementation Guide v1.2	Defines the ADSL and the BDS and Includes implementation examples
ADaM Structure for Occurrence Data (OCCDS) v.1.0 **OCCDS Implementation Guide v1.1**	Defines the OCCDS data standard and includes implementation examples (e.g. AE, CM)
ADaM Controlled Terminology	Includes codelist definition (e.g., DATEFL, DTYPE)
Analysis Results Metadata (ARM) v1.0	Includes general specifications for analysis results metadata
ADaM Conformance Rules v2.0	Includes description of ADaM conformance rules table
ADaM Examples in Commonly Used Statistical Analysis Methods	Includes example BDS datasets and documentation that support other common analyses (e.g., ANCOVA)
ADaM Basic Data Structure for Time-to-Event Analyses v1.0	Includes example BDS datasets and documentation that support time-to-event analyse

Introduction to ADaM: Documents

- **Documents from Regulatory Agency**

FDA

Data Standards Catalog v10.4

Includes what standards and standards versions are being accepted (e.g., SDTM, ADaM, Define)

Use	Standard	Exchange Format	SDO	Property	FDA Center(s)	Date Support Begins	Date Support Ends	Date Requirement Begins [10] [11]	Date Requirement Ends	Statutory, Regulatory, or Guidance Authority Sources	Information Sources
Clinical study datasets	**ADaM**	**XPT**	**CDISC**	**ADaMv2.1**	**CDER, CBER**	**Ongoing**		**12/17/2016 [1] 12/17/2017 [2]**		**Standardized Study Data**	**CDISC.org - ADaM**
Clinical study datasets	ADaM	XPT	CDISC	ADaMIGv1.0	CDER, CBER	Ongoing	03/15/2019 [1] [12] 03/15/2020 [2] [12]	12/17/2016 [1] 12/17/2017 [2]	03/15/2019 [1] [12] 03/15/2020 [2] [12]	Standardized Study Data	CDISC.org - ADaM
Clinical study datasets	ADaM	XPT	CDISC	ADaMIGv1.1	CDER, CBER	2017-10-02		03/15/2019 [1] 03/15/2020 [2]		Standardized Study Data	CDISC.org - ADaM
Clinical study datasets	ADaM	XPT	CDISC	ADaMIGv1.2 ADaMIGv1.3	CDER, CBER	07-18-2022		03-15-2024		Standardized Study Data	CDISC.org - ADaM
Clinical study datasets	**SDTM**	**XPT**	**CDISC**	**SDTMv1.1**	**CDER, CBER**	**Ongoing**	**01/28/2015 [12]**			**Standardized Study Data**	**CDISC.org - SDTM**
Clinical study datasets	SDTM	XPT	CDISC	SDTMIGv3.1.1	CDER, CBER	Ongoing	01/28/2015 [12]			Standardized Study Data	CDISC.org - SDTM

Study Data Technical Conformance Guide v5.8

Includes practical submission guidance (e.g., dataset size, variable name and dataset label length)

PMDA

PMDA has its own data standards catalog and technical conformance guide

- Conformance may differ from FDA

Fundamental Principle

1 An ADaM datasets must support traceability

2 An ADaM datasets must be accompanied by metadata

3 An ADaM datasets must be usable by commonly available software tools

4 An ADaM datasets must be analysis-ready

- Is this Analysis Ready ?

```
data Age_Table;
  set adam.adsl;
  If age>50 then agegr1='>50 yrs';
  Else If age ne . then agegr1='<=50 yrs';
run;
proc freq data=Age_Table;
  Table trt01p*agegr1;
run;
```

Fundamental Principle – Traceability (1/2)

What is Traceability?

- Traceability is the property that enables the understanding of the data's lineage and/or the relationship between an element and its predecessor(s).
- Traceability permits the understanding of the relationship between the analysis results, the analysis datasets, and the SDTM datasets
- Traceability establishes across-dataset relationships as well as within-dataset relationships
  - In regulatory reviewer perspective, traceability may be even more important than data standards

Type of Traceability?

- Metadata Traceability:
  - Provided by Define-XML files via standardized metadata for analysis datasets, analysis variables, analysis parameter vale-level metadata and analysis result metadata (Mandatory for PMDA, not FDA)
- Data point Traceability:
  - Provided by specific variables/records in ADaM datasets which help to understand where a value came from (e.g., SRCDOM, SRCSEQ, SRCVAR)

Fundamental Principle – Traceability (2/2)

- The ASEQ variable can be used to provide traceability.

Traceability to ADaM

Traceability back to SDTM / other ADaM

USUBJID	ASEQ	AVAL	SRCDOM	SRCSEQ	SRCVAR	PARAMTYP
XYZ-1001	1	8	QS	10	QSSTRESN	
XYZ-1001	2	4	QS	11	QSSTRESN	
XYZ-1001	3	7	FA	2	FASTRESN	
XYZ-1001	4	16	ADXX	1	AVAL	DERIVED
XYZ-1001	5	8	ADXX	2	AVAL	DERIVED
XYZ-1001	6	14	ADXX	3	AVAL	DERIVED

※출처: https://pharmasug.org/proceedings/2013/PO/PharmaSUG-2013-PO13.pdf

Misunderstandings about ADaM Implementation (1/2)

- **Overuse of Traceability**

1

USUBJID	PARAM	AVAL	SRCDOM	SRCSEQ	SRCVAR
ZZZ101-01	Total Cholesterol (mg/dL)	210	LB	19	LBSTRESN
ZZZ101-02	Total Cholesterol (mg/dL)	178	LB	21	LBSTRESN

2

USUBJID	PARAM	AVAL	LBSEQ
ZZZ101-01	Total Cholesterol (mg/dL)	210	19
ZZZ101-02	Total Cholesterol (mg/dL)	178	21

※출처: Common Misunderstandings about ADaM Implementation

Misunderstandings about ADaM Implementation (2/2)

- **What if AVAL is derived from a character version of laboratory results?**

> **[Statistical Analysis Plan]**
>
> *…All numeric values recorded BLQ or above the upper limit of quantification are set to their respective limits for all summaries.*

1

PARAM	PARAMN	AVAL	AVALC
C Reactive Protien	1	0.02	<0.02
C Reactive Protien	1	0.02	0.02

PARAM	PARAMN	AVAL	AVALC	LBSTRESN	LBSTRESC
C Reactive Protien	1	0.02	<0.02		<0.02
C Reactive Protien	1	0.02	0.02	0.02	0.02

2

PARAM	PARAMN	AVAL	AVALC	LBSTRESN	LBSTRESC	DTYPE
C Reactive Protien	1	0.02			<0.02	LLoQ
C Reactive Protien	1	0.02		0.02	0.02	

Technical Considerations (1/2)

CELLTRION

1 File Naming [1]

- Prefer lowercase characters.
- Avoid using spaces and underscores; use a hyphen instead.
- Do not use any of these common illegal characters or symbols.

2 Do not change the dataset file name after finalization

- Internal dataset name should be match the file name.

3 SDTM/ADaM dataset [2]

- Only single byte data is accepted.
- Data which will be submitted has to be English (single byte).

[1] https://www.cdisc.org/documentation/file-naming
[2] Japanese submission/approval processes from programming perspective

Technical Considerations (2/2)

4 Validation – Pinnacle 21

- Pinnacle 21 Community is a free, open-source tool for assessing standards compliance.
- Pinnacle 21 Enterprise is designed to ease regulatory submission preparation, manage standards for datasets, and provide continuous compliance.
- Pinnacle 21 Enterprise is used by regulatory agencies (FDA and PMDA [3]) for validation of dataset.

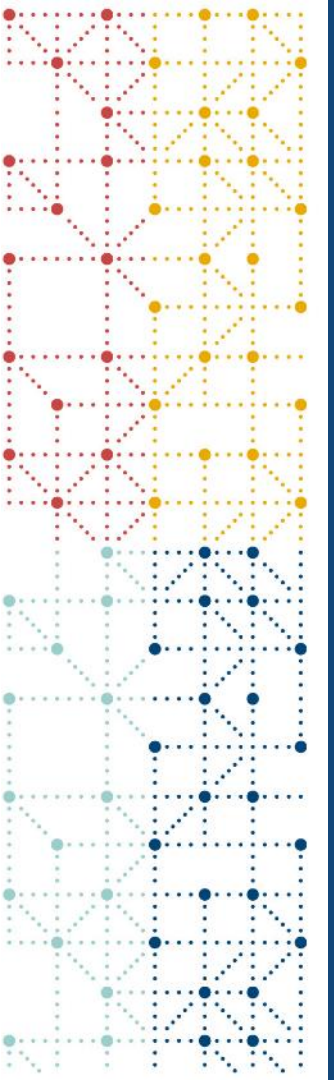
5 ADRG - Issues Summary

- Summarize findings from an ADaM conformance report in table form.

6 Submit a Standardized Data Sample to FDA

- The FDA offers a process for submitting sample standardized datasets for validation [4].

[3] https://www.pmda.go.jp/english/review-services/reviews/0002.html
[4] https://www.fda.gov/drugs/electronic-regulatory-submission-and-review/submit-standardized-data-sample-fda



CDISC submission case studies

Insufficient Quality

Insufficient Quality

- What is the worst outcome to ourselves?

The quality of the original data submission was not optimal.

Data quality related issues were identified throughout the review process.

21st Century Review timelines were not met for the primary clinical and statistical reviews as a result of the data quality and integrity. The applicant withdrew the priority request voluntarily. The review clock had to be extended from 6 months to 10 months.

Insufficient Quality – Cases (1/5)

- What went wrong?

Company A

Company B

Company C

Company D

Company E

3.1 Submission Quality and Integrity

The quality of the submission did not permit an efficient and timely review. The key determinants of this assessment included the following:

- Missing components of the eCTD in the original NDA submission
- Data discrepancies resulting from, at least in part, inconsistent data cutoff dates used to generate the datasets and to create the patient case report forms (CRF) in the NDA submission. As a result, all safety findings based on the ISS database required reanalysis using the ISS database that was submitted in the 120-Day safety update.
- Data discrepancies resulting from errors in the eCTD submission documents
- Dataset definition files which did not contain definitions of multiple variables
- Dataset definition files which contained an inadequate level of detail in the variable definitions to facilitate efficient review
- Inadequate and/or incorrect annotations within the annotated CRFs
- Key variables in datasets were absent, inconsistent in name or definition across datasets, and/or incomplete
- Non-functioning SAS programs for statistical analyses

Insufficient Quality – Cases (2/5)

- What went wrong?

Company A

Company B

Company C

Company D

Company E

2.2 Data Sources

All data sources are included in the sponsor's eCTD submission located in the FDA/CBER Electronic Data Room (EDR).

Reviewer Comment: [redacted] submitted an amendment on 23 April 2013 providing additional information on datasets and programs for the review of the study results. The amendment states that the inhibitor test result itself is stored in the value AVAL. This reviewer has discussed with the clinical reviewer that "AVAL" is a typographical error and that the variable AVALC stored the value for the inhibitor result.

Insufficient Quality – Cases (3/5)

- What went wrong?

Company A

Company B

Company C

Company D

Company E

3.1 The quality of the Submission

The reviewer is able to reproduce the primary endpoint, the HIV RNA level, from the raw dataset. On the other hand, the sponsor's define file can be defined better to facilitate the regulatory review. For example, in the define file for the raw data, the important variable "QUALIF" (the censoring indicator of the HIV RNA level) from the HIV RNA dataset has no definition. We finally understood the importance of the variable but it took several rounds of communications with the sponsor before we were able to trace to this variable and understand its importance. For another example, on the [redacted] (the dataset of study medication) page of the file blackcrf_raw.pdf, it is stated "Refer to Appendix 3" but Appendix 3 was not submitted until we requested after we realize it is missing. These types of problems exist abundantly in the submission and created unnecessarily extra work.

Insufficient Quality – Cases (4/5)

- What went wrong?

Company A

Company B

Company C

Company D

Company E

3.1 Data and Analysis Quality

There seemed to be two ways of programming to get the efficacy result after reading the data definition files, and the reviewer's guide, however, we got two different p-values, with one smaller than 0.05 and the other larger than 0.05, but neither of them was the same as what the sponsor presented in the study report.

...

We were also not able to understand the patient level flags in the ADEFF data file. The efficacy table presented in the study report has the LOCF already applied, we could not separate the number of patients whose efficacy result were imputed and the details on the missing pattern of these patients, such as which time point (10 mins or 90 mins) exactly was used to carry forward for the 2-hour value. So, our IR also included questions to clarify exactly how many patients have observed (not imputed) values at each time point post dose for each group and asked the sponsor to provide more details on the LOCF imputation method.

Insufficient Quality – Cases (5/5)

- What went wrong?

Company A
Company B
Company C
Company D
Company E

3.1 Data and Analysis Quality

There are some issues of data and analysis quality, but none of them is considered a major issue. The issues are: poor documented traceability of derived variables from tabulation data sets, poor documented traceability of tabulation data sets from case report forms, and complex and confusing language used in the Case Report Form for the two efficacy assessments of clinical cure and recurrence.

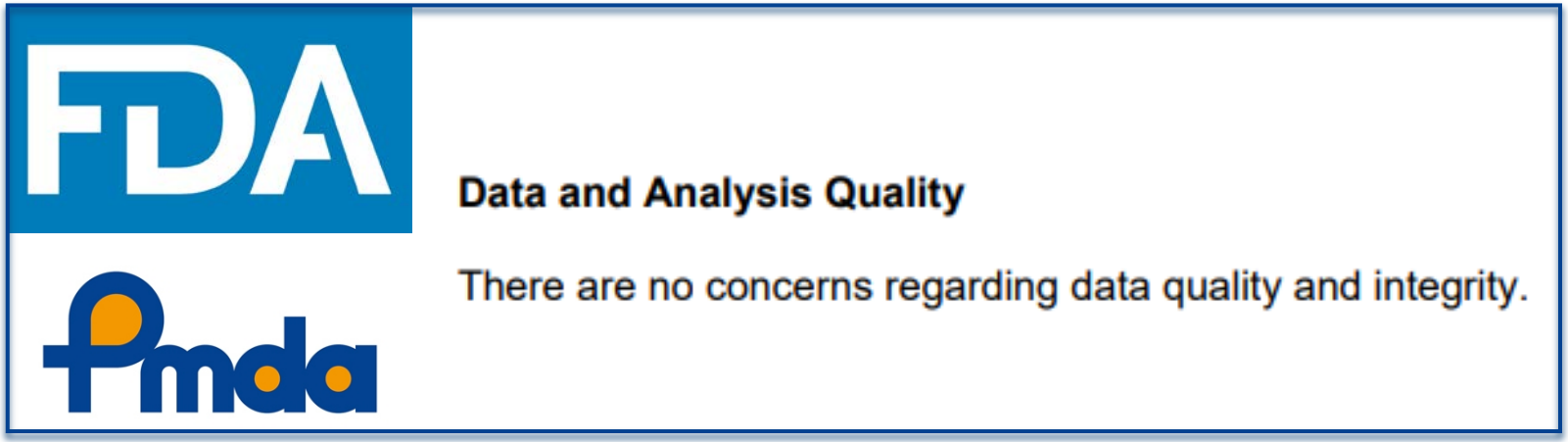
There is minimal derivation involved in the primary and two secondary endpoints. Clinical Cure and Recurrence are the two main efficacy outcomes, they are clinically reported outcomes directly captured in a CRF. These two assessments had very little missing values (1 missing value for Clinical Cure and 15 missing values for recurrence among cured) which were set to failures as pre-planned in protocol. Thus, these two assessments matched the data entry in the CRF most of the time and with minimal imputation of failures for missing values. There is minimal derivation for the endpoint of global cure, since it is a composite of clinical cure AND no-recurrence during follow up.

Summary and Suggestions

Key lessons-learned from our own experience

Our Goals and Values

- **Sufficient Quality**



Thank You!