



Sumptuous Data Sciences

DATA COLLECTION TIME POINT – CHALLENGES WITH ANALYSIS AND SUBMISSION

Aayush Bhoraskar
Sumptuous Data Sciences
Toronto, Canada



Sumptuous Data Sciences

Agenda

- Introduction
- Types/Instances of Data Collection
- Different data collection stages/cases



Sumptuous Data Sciences

Introduction

- Clinical Programming
- E-Sub (electronic submission to FDA)

Types/Instances of Data Collection

- Interim data
 - Limited data
 - Near-perfect representation of final study data
- Following database lock
- Mapped from legacy standards to CDISC standards
- Integrated from multiple studies

Case 1: Interim Analysis, Limited Data

- Not complete representation of study population.
- Major Issues
 - Programming difficulties.
 - Missing information for CRF annotation.
 - Validation issues.
 - Further data extraction affecting *define.xml*.



Case 2: Interim Data

- A near perfect representation of the study data.
- Major issues
 - Mapping issues.
 - Programming difficulties.
 - Lot of split domains.
 - Further data extraction affecting *define.xml*.

Case 3: Data following database lock

- Completed study
 - Least missing information.
 - This is the best case scenario from a programming and submission perspective.
- Major issues
 - Data issues require unlocking, updating, and re-locking the database.
 - Inability to resolve reported validation issues.

Case 4: Mapped data from legacy

- Data is collected in legacy format, and the mapped from that format to SDTM/ADaM domains.
- Major Issues
 - Difficulty mapping entire collected information to CDISC standards.
 - CRF annotation can be challenging due to complex mapping specs.
 - Traceability issues might occur because data collected is not in CDASH format.

Case 5: Integrated data

- This implies that datasets are generated by merging information from multiple studies.
- Major Issues
 - May require mapping of all studies to a common submission standard (like CDISC).
 - Traceability issues.
 - Any issue from other data scenario considering it has multiple different types of data to merge.



Summary

- Different scenario have different issues.
- Interim data generally has programming and mapping issues.
- Mapping from legacy standards can result in several split domains or complex re-combination of data.
- Integrated data can have all these issues given data it's sources.
- Information may be explained in DRGs, but best if considered as last option.



Sumptuous Data Sciences

THANK YOU! QUESTIONS?

Aayush Bhoraskar

aayush.bhoraskar@sumptuous-ds.com