Technical Implementation of FDA Requirements

Overview of FDA Study Data Technical Conformance Guide

Sergiy Sirichenko
January 7 & 8, 2015

PINNACLE 21

Presenter – Sergiy Sirichenko

- Principal Data Scientist at Pinnacle 21
- Co-founder of OpenCDISC
- Data Fitness Analyst for FDA JumpStart and eData Teams
- Validation rule developer for FDA DataFit and OpenCDISC
- The person who answers all your questions on OpenCDISC Forum

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 Food and Drug Administration.

Abbreviations

- SDTCG – FDA Study Data Technical Conformance Guide
- IG – Implementation Guide (e.g., SDTM IG)
- SDRG – Study Data Reviewers Guide
- ADRG – Analysis Data Reviewers Guide

Topics

- Technical part of SDTCG
- Content of Chapters 3-8
- Details which are important or new
- Examples of common issues

#1.2 Purpose

"This Guide provides technical recommendations to sponsors for the submission of … study data and related information in a standardized electronic format … The Guide is intended to complement and promote interactions between sponsors and FDA review divisions. However, it is not intended to replace the need for sponsors to communicate directly with review divisions regarding implementation approaches or issues relating to data standards…"

P21

#3 Exchange Format

- XML
- PDF
  - Documents
  - Supported version in Standard Catalog
- File Transport Format
  - SAS XPORT v5, not SAS CPORT
  - One dataset per file
  - Dataset name the same as XPT file
    - Common issue in non standardized data like PK, PG

- Dataset Size
  - Dataset > 1 GB should be split in smaller files
  - Submit both non-split and split datasets
  - “split” folder (see section #7)
- Dataset Column Length
  - Maximum length of variable used
- Variable and Dataset Descriptor Length
  - SAS limitations:
    - Variable Name - 8 Chars
    - Variable and Dataset Label - 40 Chars

- Special Characters: Variables and Datasets
  - ASCII text codes only
- Variable and Dataset Names
  - No punctuation, dashes, spaces or other non-alphanumeric symbols
- Variable and Dataset Labels
  - May include punctuation characters
  - No special characters like
    - Unbalanced apostrophe, quotation marks, parentheses, braces, brackets
    - “<“ and “>” signs
    - “Parkinson’s”

#4 Study Data Submission Format

- #4.1 CDISC
  - *Standards Catalog* provides a listing of supported data standards and versions
  - Analysis files are critical
  - ADaM is expected for SDTM data
  - Currently ADaM specifications for SEND have not been developed
  - When IGs do not provide specific instructions on certain study data, Sponsor should discuss with FDA

P21

#4.1.1 SDTM

"It is recommended that sponsors implement the SDTM standard for representation of clinical trial tabulation data prior to the conduct of the study. The use of case report forms that incorporate SDTM standard data elements (e.g., Clinical Data Acquisition Standards Harmonization (CDASH)) allows for a simplified process for the creation of SDTM domains."

P21

"If there is uncertainty regarding implementation, the sponsor should discuss application-specific questions with the review division and general standards implementation questions with the specific center resources identified elsewhere in this Guide (See section 1.2). When data imputation is utilized, sponsors should submit imputed data in an analysis dataset, and the relevant supporting documentation (e.g., ADRG, define.xml) explaining the imputation methods."

"Except for variables that are defined in the SDTMIG as being coded, no numerically coded variables should typically be submitted as part of the SDTM datasets"

- SUBJID
  - ID of the entity (i.e., person) in trial
  - If the same subject is screened more than once, then SUBJID should be different
- USUBJID
  - Unique across the entire application
  - The same USUBJID across all datasets
    - Common issue – PK, PG, SDTM/ADaM, ISS
  - No leading or trailing spaces
  - No inconsistency in usage “0” (S01-001 vs. S1-1)
  - Improper implementation may results in request for Sponsors to re-submit their data

- Adjudication Data
  - There are no existing standard or best practice
  - Advises Sponsors to
    - discuss their approach with review division
    - include details in SDRG on
      - presence
      - implementation approach
      - location

#4.1.1.3 SDTM Domain Specifications

- SUPPQUAL
  - Should be used only for key data which does not fit in SDTM domains
    - Examples of common issues:
      - SUPPQUALs keep all EDC variables
      - 300+ SUPPQUAL variables
  - Discuss with review division, document in SDRG

- DM (Demographics)
  - Single record per subject
  - ARM is blank for Screen Failures
    - “SCRNFAIL” -> “”
  - ACTARM is blank for Not Treated subjects
    - “NOTTRT” -> “”
- DS (Disposition)
  - EPOCH should be populated for all records
  - DEATH should be the last subject record

- SE (Subject Elements)
  - Should be included
- AE (Adverse Events)
  - Treatment Emergent Flag is expected
  - All AEs from CRF should be included
  - AE Seriousness Criteria should be provided. This info is critical
- Custom domains
  - Confirm that there are no standard domains
    - Check recent versions of IG
  - Provide details in SDRG

- LB (Lab Test Results)
  - Large size is a common issue
  - Split to < 1GB files according to LBCAT and LBSCAT if needed
- Trial Design (TS, TA, TE, TV, TI)
  - Should be included

P21

#4.1.2 Analysis Data Model

"Generally, ADaM analysis datasets facilitate FDA review. However, it does not always provide data structured in a way that supports all of the analyses that should be submitted for review. For example, ADaM does not support simultaneous analysis of multiple dependent variables or correlation analysis across several response variables. Therefore, sponsors should, as needed, supplement their ADaM datasets after discussions with the specific review division."

P21

- Key Efficacy and Safety Variables
  - Efficacy datasets are expected
  - Detailed documentation is needed
- Timing Variables
  - In addition to protocol-scheduled visit variable, at least two additional timing variables are expected
    - AVISIT, AVISITN
- Core Variables
  - Should be listed after key variables
- Dates
  - Numeric
  - In addition to ISO8601
  - Imputation, Flag, documentation

- Labels
  - Should be unique and different from SDTM
    - "Adverse Events" is not correct label for ADAE
- Software Programs
  - Provide programs used to create ADaM datasets and TFLs
  - ASCII text or PDF files
  - File names with reference to software
    - "adae.sas.txt"
  - Sufficient documentation

- ADSL (Analysis Data Subject Level)
  - Required
  - Important study specific baseline subject characteristics and covariates presented in protocol
- Imputed data
  - When data is imputed, it should be submitted in analysis datasets
  - Detailed relevant supporting documentation
    - Define.xml
    - ADRG
    - Complicated algorithms, etc.

#4.1.3.3 SEND

- Similar to SDTM
- MIORRES -> MISTRES modifiers
- MA domain expect to use VISITDY
- tumor.xpt is expected for oncology studies

P21

#4.1.4 General Considerations

"For the purposes of SDTM and SEND submissions, all Required, Expected, and Permissible variables that were collected, plus any variables that are needed to compute derivations, should be submitted.

SDTM datasets should not contain imputed data. FDA recognizes that SDTM contains certain operationally derived variables that have standard derivations across all studies (e.g., --STDY, EPOCH). If the data needed to derive these variables are missing, then these variables cannot be derived and the values should be null."

Examples of FDA expected variables

- Baseline flags
  - LB, VS, EG, PC, MB
  - If data were collected or can be derived
- EPOCH
- Study Days
  - When --DTC, --STDTC, --ENDTC collected, then populate --DY, --STDY, --ENDY

#4.1.4.5 Data Definition File

- "define.xml"
- *"A properly functioning define.xml file is an important part of the submission of electronic study datasets."*
- define.pdf is also expected for define.xml v1.0
- Recommended to send test files to FDA eData team prior to submission
- An insufficiently documented define file is a common deficiency
- Stylesheet files for define.xml are required

#4.1.4.6 Annotated Case Report Forms

- New name "*acrf.pdf*" instead of "*blankcrf.pdf*"
- Mapping of each field on CRF to dataset variable
- aCRF should include variable *names* and *coding* for each CRF item
- If some data are recorded on CRF but not submitted
  - annotate with "*NOT SUBMITTED*" text
  - explain in SDRG

#5 Therapeutic Area Standards

"This section is reserved for future comments, recommendations, and preferences on therapeutic area data standards."

P21

#6 Terminology

- *"Common dictionaries should be used across all clinical studies and throughout the submission for each of the following:"*
  - AE, CM, PR, MH, indications and drug names
- See *Standards Catalog for* recommended usage of terminology
- Conformance with Standard Terminology is required
- Common issues:
  - Misspelling
  - Not following upper/lower case
  - Use of hyphens

- Use of Control Terminology
  - *"use the most current version of an FDA-supported terminology available at the time of coding"*
  - Different studies may use different versions
  - Impact of usage of old versions should be described in SDRG and Standardization Plan
  - Pooled analysis (e.g., ISS) must use a single version of terminology
- Maintenance of Controlled Terminology
  - *"good terminology management practice"*
  - Creation of custom terms is discouraged
  - Consistency throughout the application
  - Standardization Plan, SDRG

P21

Adverse Events

- MedDRA
  - Exact spelling and capitalization
  - Single version for ISS

Medications

- FDA Unique Ingredient Identifier (UNII)
  - TS domain, TSPARMCD=TRT, COMPTRT, CURTRT, …
- WHO Drug Dictionary
  - CMDECOD – generic name
  - CMCLAS – class or ATC level 4
  - ATC codes in SUPPCM

P21

Pharmacologic Class

- National Drug File – Reference Terminology (NDF-RT)
  - TS domain, TSPARMCD=PCLAS

Indication

- SNOMED CT
  - TS domain, TSPARMCD=INDIC, TDIGRP
  - Harmonization with Structured Product Labeling (SPL)

#7 Electronic Submission Format

- eCTD
- define.xml and supportive stylesheets in the same folder as datasets
- No empty folders
- New "*misc*" folder is introduced instead of "*listings*"
- For need of additional folders consult with FDA

P21

#8 Data Validation and Traceability

"data validation is a process that attempts to ensure that submitted data are both compliant and useful. Compliant means the data conform to the applicable and required data standards. Useful means that the data support the intended use (i.e., regulatory review and analysis). "

- Study Data Validation
  - Conformance validation and Quality checks
  - Links to FDA rules on Standards Web page
  - Sponsors should fix issues and explain in SDRG why certain errors could not be corrected

- Study Data Traceability
  - Important component of regulatory review
  - Relationship between
    - Analysis results
    - Analysis datasets
    - Tabulation datasets
    - Source data
  - Standards are helpful (CDASH)
  - Standardized data will be required
  - Traceability issues with Legacy data conversion

Get updates about our events

 linkedin.com/company/pinnacle-21

 twitter.com/pinnacle_21

 facebook.com/pinnacle21

 pinnacle21.net/blog

 opencdisc.org

Questions

Sergiy Sirichenko
ssirichenko@pinnacle21.net