The CDISC/FDA Integrated Data Pilot: A Case Study in Implementing CDISC Standards to Support an Integrated Review

d-Wise Technologies
Chris Decker
Life Sciences Director

Overview

- Pilot Mission and Goals
- Summary of the Pilot Data
- Analysis Methodology
- SDTM Findings
- ADaM Findings
- Integrated Findings
- Define.xml Updates
- Deliverables, Pilot Status, and Timelines

Team Members – Thank you!

Co-Leaders: Chris Decker, Steve Hirschfeld, Yuguang Zhao

Greg Steffens	Robert Collins	FDA:
Nate Friemark	Asif Taiyabi	Suzanne Demko
Adam Huang	Karen Wade	Joyce Korvick
Aileen Yam	Thaddea Dreyer	Thomas Marciniak
Eric Qui	Brian Mitchell	Christine Nguyen
Pam Cahalin	Geoff Mann	Nancy Snow
Sandy Lei	Ann Krebs	Ana Szarfman
Ian Fleming	Arline Nakanishi	Gwen Zornberg
Ramana Setty	Ugochi Emeribe	Mina Hohlen
Archie Chaudhari	Richard Lewis	Patty Garvey
Deven Dharm	Lex Jansen	Amy Malla (CBER)
Bob Dainton	Carrie Mazzrillo	Crystal Allard (CBER)
Frank Roediger		Jingyee Kou (CBER)
		Gale Heavner (CBER)
		Jessica Kim (CBER)

Integrated Data Pilot Mission

The mission of the CDISC Integrated Data Pilot is to demonstrate that a data submission created using CDISC Harmonized Standards will meet the needs and expectations of FDA reviewers in conducting an **integrated** review of data from **multiple** studies and compounds.

CDISC Pilot Goals

- Assess the applicability of the CDISC models to integrate data
- Identify any issues/questions to be addressed by the CDISC Teams
- Validate the components of the CDISC models can effectively be used together
- Evaluate the most current CDISC models including SDTM, ADaM, ODM/Define.xml, and Trial Design

CDISC Pilot Goals

- Use the CDISC models to support integrated standard analysis and reporting as described within relevant FDA Guidance
- Support the critical path initiatives around standard data, integrated databases, standard data collection, and studies of special populations
- Anything that Becky or Dave asks us to do

Overview of Pilot Data

- 3 compounds/8 studies
  - Multiple Clinical trials
  - Population included children with hypertension
- Trial Designs
  - Pharmacokinetics trials
  - Randomized double-blind trials
- Artificial test data composed of de-identified and randomly modified data; provided as SDTM practice datasets

Overview of Pilot Data

- Documentation provided:
  - One page summary of each study including study design schedule
  - De-identified data sets
- Initial domains provided: DM, AE, VS, EX, CM, DS, SC, PC, LB, MH

Analysis Approach

- Different then conventional analysis
- Limited by availability of documentation
- Decided to provide a integrated review of patient experience
  - Not an ‘ISS’
  - Not a comparison of Drug vs. Placebo
  - Abnormal ‘Events’ compared to Exposure

Analysis Approach

- Events include both adverse events, abnormal laboratory values, and abnormal vital signs
- Final abbreviated reports with descriptive statistics only
- Used standard reference ranges for vital signs and laboratory values
- All summaries will use ADaM domains
- Events Overall and by exposure, age, and gender

Pilot Framework

Defined three groups:

- Analysis and Design – analysis plans, ADaM requirements, and Study Reports
- Programming – managed metadata and programmed data sets and summary tables
- Package – organized components, extended define.xml and rendered the final views

Metadata Framework

- Metadata collected in Excel spreadsheets
- Converted to SAS data sets and used for validation, adding attributes, and define

	A	C	G	H	I	J	K	L	
1	table	column	clabel	clabellong	ctype	clength	cformat	cformatflag	computation_method
260	ADLB	LBCAT	Category for Lab Test		C	15			See row level metadata
261	ADLB	LBSEQ	Sequence Number		N	8			Equal to LB.LBSEQ. Any new records will be missing.
262	ADLB	PARAM	Parameter Description		C	200			LB.LBTEST || " (" || LB.LBSTRESU || ")"
263	ADLB	PARAMCD	Parameter Code		C	8	LBPARACD	2	Equal to LB.LBTESTCD for all records carried over from LB

table	column	param	paramrel	ptype	plength	pformat	pformatfla	pdescription	computation_method
ADLB	PARAMCD	ALB	AVAL	N	8	4.1	1	ADLB_AVAL	Equal to LB.LBSTRESN
ADLB	PARAMCD	ALB	LBCAT	C	20	LBCATBC	2	ADLB_LBCAT	Equal to LB.LBCAT
ADLB	PARAMCD	ALB	PARAM	C	20	LBTALB	2	ADLB_PARAM	Equal to LB.LBTEST Plus units MM

SDTM Development

- Data was supplied in SDTM ‘like’ format
- Still took 3 months to clean it up
- Most of the issues were just bad data – typical legacy data issues
  - Interpretation of SDTM…
  - Derived Data
  - Baseline Flags

ADaM Development

- Cross team members made communication easier
- Used the current draft ADaM Implementation Guide
- Model open to interpretation – significant discussions back and forth
- Implementation Guide is still ‘draft’
- 6 months to hash through all these changes

ADaM Development

- Defining analysis subgroups
  - robust and useful for everyone (ranges, variables, numeric/character)
  - Defined three analysis subgroups – both character and numeric
  - Followed naming conventions in ADaM IG
- Treatment variables – How to implement TRTxP variables for the more complex multi phase studies
  - Include TRTxP only if it used in the analysis? We included everything
- Defining Phases within ADaM
  - Analysis Phases/Visits not the same as SDTM (EPOCH/VISIT)
  - Naming conventions? APHASE/AEPOCH?

ADaM Development

- Relationship between SDTM and ADaM
  - What variables should you copy from SDTM to ADaM? Used in Analysis? Supportive? We only copied what was used in analysis
  - If used in Analysis should we rename, 'A' prefix? We renamed.
- Analysis Flags within ADSL
  - ADSL Flags – Confusion over flags vs. subgroups
  - Defined multiple flags for abnormal AEs, Vital Signs, and Labs
- Record Level Flags
  - Analyzed records or records that could be analyzed?
  - Second flag to flag records that met a certain criteria. Defined as flag but should have probably used the Criteria rules

ADaM Development - Integration

- Building integrated ADaM using study ADaM
- Proof of concept – defined two major changes
  - Exposure quartiles
  - Definition of EPOCH across studies
- Challenges in Integrating different study designs
  - Led to defining integration for similar designs
  - Other issues identified (e.g. time variables)

Integrating ADaM

- Having Standardized ADaM made it easier
- However, still need to address differences
  - Study designs
  - Differences in data collected
- Integrating the standard was more efficient but still required additional work

Define.xml Tasks

- Worked with define.xml team to provide support for row level metadata
- Adding extension for Analysis Results Metadata based on Pilot 1 findings
- New style sheets and a PDF rendering of the define.xml for human readable format

Define.xml Value Level

- Explored multiple ways to support metadata
- First method involved extending define
  - Seemed to work
  - Ran into challenges capturing value level methods
- Redoing the implementation based on define team feedback

Variable	Label	Key	Type	Length	Code List / Controlled Terms	Derivation	Origin	Role
USUBJID	Unique Subject Identifier	1	text	22				Identifier
VISITNUM	Visit Number	2	float	8				Timing
LBDTC	Date/Time of Specimen Collection	3	text	19				Timing
LBTESTCD	LAB Test or Examination Short Name	4	text	8	LBTESTCD	LB_LBTESTCD		Topic

Parameter Value	Parameter Related Variable	Type	Length	Format	Code List / Controlled Terms	Derivation
Context: In LB where LBTESTCD is ALB	LBCAT	text	20		LBCATBC	LB_LBCAT_PV

LBCATBC, Reference Name (CODELISTC14)
Valid Values
BLOOD CHEMISTRY

LB_LBCAT_PV	Computation	Equal to LBNORMALS.LBCAT merged on LBTESTCD

Define.xml – Analysis Results

- Capturing metadata for analysis results
- Added in first Pilot as an one off extension
- Worked with Define to define extension

Analysis Results

Display Identifier	Display Name	Result Identifier	Param	ParamCD	Analysis Variable	Reason	Dataset	Selection Criteria	Documentation	Model Specification
Table 14-3.01	Primary Endpoint Analysis:ADAS Cog(11) - Change from Baseine to Week 24 - LOCF	Analysis of dose response	ADAS-Cog (11) Total Score	ACTOT11	CHG	Pre-specified in Protocol	ADQSADAS	EFFFL='Y' and ITTRFL='Y' and AVISIT='Week 24' and PARAMCD='ACTOT11'	SAP_SEC_10.1.1, SAP_TEMP_5. Linear model analysis of dose response for the ADAS-Cog(11) total score change from baseline at Week 24 - missing values imputed using LOCF, Efficacy population. Used PROC GLM in SAS to produce p-value (from Type III SS for treatment dose); Independent terms in model are TRTDOSE (randomized dose: 0 for placebo; 54 for low dose; 81 for high dose) SITEGRP (site group, as a class variable) and BASE (baseline ADAS-Cog score).	PROC MIXED; CLASS USUBJID SITEGRP AWEEK TRT1P; MODEL CHG = TRT1P SITEGRP AWEEK TRT1P*AWEEK BASE BASE*AWEEK / OUTP=PRED DDFM=KR; REPEATED AWEEK / SUBJECT=USUBJID TYPE=UN; LSMEANS TRT1P / DIFF CL; RUN;

Package Challenges

- Separate define files for SDTM and ADaM?
  - Issues with the size of the files in first Pilot
- No real good solution for linking SDTM to ADaM or integrated data to the individual study data
- Storage of style sheets and define files

Final Deliverables

- Individual Studies
  - SDTM domains and ADaM domains for 4 studies (NEW)
  - Summary tables and Study Report for studies 1/2
  - Define files for SDTM and ADaM/Analysis Results
- Integrated Data
  - ADaM domains for similar study designs within Compound A and across Compound A
  - Clinical Study Report for Compound A
  - Define files for Integrated ADaM/Analysis Results

Project Status

- Completed first two studies include data, tables, report, and define files
- Delivered first two studies to the FDA for review
- Working on Integration of Compound A
- Integrated data and reports in December
- Project report available by the end of the year

Pilot Summary

- Volunteer projects are not simple
  - Deliverables added during the process
  - Testing a lot of different parts that are moving
- Models are open for interpretation – lead to a variety of ‘ways’ of doing it
  - Provided good feedback to CDISC teams
- ODM/Define can be extended but is challenging
- Following standards leads to many efficiencies but there is still work