Need for Human Involvement in Checking for Compliance of Both SDTM and ADaM

Octagon Research Solutions, Inc.

Octagon®

The Clinical Data and Regulatory Experts

Sandra Minjoe
ADaM Consultant
Ken Stoltzfus
CDISC Consultant

June, 2012

Automated Validation Tools

- Integral part of the validation process
- Concise, repetitive checks that increase data accuracy and reduce validation time
- Identification of the most common errors and inconsistencies with the SDTM/SDTMIG, ADaM/ADaMIG

BUT:

- The number of checks is small compared to the number of implementation decisions
- Not every SDTM/SDTMIG convention can be represented in computer-executable code
- May not provide clear explanations of how to fix errors such as source-data vs. data-conversion issues
- The creativity of humans will always outpace the development of checks to limit that creativity

Automated Validation Tools

- SDTM and ADaM each publish a list of automatable compliance checks
  - Available to CDISC members in members-only section of website
- Validation tools that do compliance checks include
  - OpenCDISC
  - WebSDM
  - Custom in-house systems
- Additional automatable checks can also be written

SDTM Compliance Checks

- Some areas where Humans are needed
  - Domain Content
  - Supplemental Qualifiers
  - Relationships
  - Trial Design

SDTM Domain Content

- Mapping to the Right Domain
  - Same data mapped to different domains (e.g., PD)
  - Unlike data have been migrated into a single domain to minimize number of custom domains
- Misuse of standard variables (hijacking)
  - --CAT and --SUBCAT
  - --GRPID
  - Time-point variables

SDTM Supplemental Qualifiers

- Problem areas include the following:
  - Analysis data
  - Additional representations of the same data
    - Numeric and character codes
    - Legacy controlled terminology
    - Separate dates and times
    - Numeric SAS dates
  - Non-unique QNAMs for relating to the same parent record

SDTM Relationships

- Documenting relationships can be especially important:
  - Findings About, when the Event or Intervention Record is submitted
  - In PK and PD domains to relate sets of parameters to specific time-effect or time-concentration curves
  - In MB / MS domains to link organisms to their susceptibility results

- Automated tools are not capable of recognizing a missing RELREC relationship
  - Record-to-record
  - Dataset-to-dataset

Example: Relating Microbiology Data

Microbiology Specimen (MB) Dataset

STUDYID	DOMAIN	USUBJID	MBSPID	MBTESTCD	MBORRES
2007-001	MB	ABC-0001	ORG01	ORGANISM	Streptococcus pneumoniae
2007-001	MB	ABC-0001	ORG02	ORGANISM	Klebsiella pneumoniae

Microbiology Susceptibility Test (MS) Dataset

STUDYID	DOMAIN	USUBJID	MSSPID	MSTEST	MSORRES	MSORRESU
2007-001	MS	ABC-0001	ORG01	Penicillin	0.25	ug/mL
2007-001	MS	ABC-0001	ORG01	Streptomycin	0.1	ug/mL
2007-001	MS	ABC-0001	ORG01	Erythromycin	0.5	ug/mL
2007-001	MS	ABC-0001	ORG02	Penicillin	0.4	ug/mL
2007-001	MS	ABC-0001	ORG02	Streptomycin	0.02	ug/mL
2007-001	MS	ABC-0001	ORG02	Erythromycin	0.25	ug/mL

SDTM Trial Design

- An accurate representation of a trial requires an understanding of:
  - the protocol
  - the CRF
  - the collected data
  - the SDTM/SDTMIG
- Tools can check
  - All necessary tables and SDTM Required variables are present
- Tools can't determine when
  - TA, TE, and TV are an inaccurate representation of the protocol
  - SE and SV were created incorrectly

SDTM Case Study

Background

- Octagon received SDTM datasets from client
- Scope of work was to review for compliance to SDTM

Compliance Issues

- OpenCDISC found many issues directly
  - value for RACE not found in user-defined Race codelist
- Octagon CheckPoint (in-house tool) uncovered more issues
  - --STRESN needs to be populated each time that --STRESC is in numeric format
- Human manual review uncovered additional issues
  - VISIT and VISITNUM captured in supplemental qualifier dataset
  - Some lab tests in a custom domain rather than LB

ADaM Compliance Checks

- Some areas where Humans are needed
  - Conditionally Required variables
  - ADSL study-specific variables
  - BDS analysis-specific variables
  - Structures other than ADSL and BDS
  - Permissible variables needed for Traceability

ADaM Conditionally Required

- ADaM's "Conditionally Required" variables are difficult to automate

- Example: Was the study randomized?
  - If so, RANDDT must be included
  - If not, there should be no RANDDT variable

ADaM ADSL Compliance

- Many variables needed in ADSL are not listed in ADaMIG section 3.1 table
- ADaMIG 1.3 states

  "ADSL contains required variables (as specified in this document) plus other subject-level variables that are important in describing a subject's experience in the trial."
- CDER Common Data Standards Issues Doc states

  "In addition to the variables specified for ADSL in the ADaM Implementation Guide, it is expected that the sponsor will include multiple additional variables representing various important baseline patient characteristics."

ADaM ADSL Compliance

- Other study-specific ADSL variables include
  - Population flags
  - Number of treatment arms
  - Cohort variables
  - Protocol deviations

- How do you automate checks for study-specific variables?

ADaM BDS Compliance

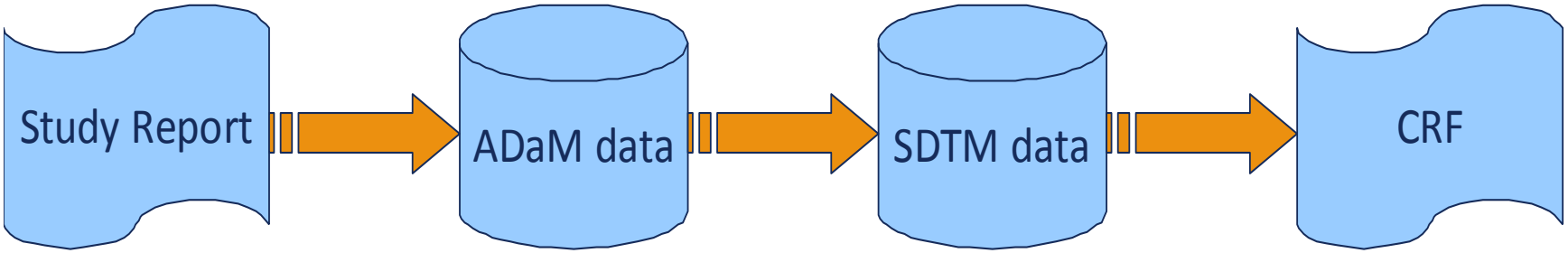
- Many BDS checks can be automated
  - Example: 1-to-1 correspondence between
    - PARAM – PARAMCD – PARAMN
- Manual checks are needed to determine
  - Were all datasets/vars/records included that are used for analysis?
  - Are DTYPE, PARAMTYP, BASETYPE used when needed, and used correctly?
  - Are any parameters included that are not used?

Other ADaM Structures

- ADaM ADAE
  - ADAE Document was finalized May, 2012
  - Not BDS structure (no AVAL/AVALC, PARAM* vars)
- Con Meds, Med Hx
  - Structured similar to ADAE for similar analysis tables
  - ADaM Occurrence sub-team is developing appendix
- Multivariate analysis data
  - Doesn't fit in BDS because need >1 analysis var on a row
  - Solution described in ADaM Additional Examples appendix
  - ADaM Multivariate sub-team is developing appendix
- Automated checks typically ignore data structures other than ADSL or BDS

ADaM Traceability

- Must be able to trace from results through datasets back to collection



- What that means for compliance checks differs across datasets and studies
  - Includes things like --SEQ, SRC*, VISIT, *DTC, and other variables copied from SDTM datasets

ADaM Case Study

Background

- Received analysis data from client
- Asked to review for ADaM compliance

Compliance Issues

- OpenCDISC found many issues directly
  - "A population variable with a suffix of FL is not present in ADSL"
- OpenCDISC found some issues that needed investigating
  - "Expected TRTSEQA is not present within dataset"
- Human review was needed to find other issues
  - Many datasets that were not in BDS structure should have been
  - ADLB contained parameters that were not analyzed

Octagon
The Clinical Data and Regulatory Experts

Summary

- Automated Checks serve a useful purpose
  - Are one part of the process
  - Good for required fields
  - Good for checks that are too time-consuming to do manually
- Automated Checks alone are not enough
  - Variability in clinical trials data
  - Virtually impossible for automated checks to identify all potential compliance issues
- Supplement Automated Checks with Manual Review
  - Conditional and permissible fields
  - SDTM domain content, relationships, trial design data
  - ADaM content for analysis needs and traceability

Future

- More information is being stored in metadata, such as
  - PRG Protocol data -> Trial Design
  - PRG SAP specs data -> ADaM
- Validation tools will likely evolve to automate even more checks

- Beware:
  - The creativity of humans will always outpace the development of checks to limit that activity

  OR

  - ***Make it foolproof and they just build a better fool***

Questions and Further Information

Sandra Minjoe
Data Standards Consulting
Tel: 1-650-255-9822
sminjoe@octagonresearch.com

Ken Stoltzfus
Clinical Data Strategies
Tel: 1-610-535-6500 (Ext. 5891)
kstoltzfus@octagonresearch.com

Acknowlegements

- Fred Wood
- Barry Cohen
- Dave Evans