CSC JumpStart – Lessons Learned

Helena Sviglin, MPH
Regulatory Information Specialist
FDA, CDER, Computational Science Center

Center for Drug Evaluation and Research
CSC
Computational Science Center

JumpStart evolved out of a desire to leverage standardized, high-quality data to give Reviewers more think-time

High quality data is the key to enabling regulatory reviewers to fully utilize the Computational Science Center's tools and services **to support decision making**

A successful CSC JumpStart Service will 'jumpstart' a review by providing the following benefits to the Review Team

	Purpose	Benefits for Reviewers
1	Assess and report on whether data is fit for purpose • Quality • Tool loading ability • Analysis ability	Understand what tools and analyses can be run and whether they might be compromised by data quality or structure issues
2	Load data into tools for reviewer use and run automated analyses that are universal or common (e.g., demographics, simple AE)	Improves the efficiency of the review by setting up tools and performing common analyses which provides the reviewer with time to focus on more complex analyses
3	Provide standard analyses to allow the reviewer to uncover safety signals that may need a focus for review	Points reviewer to a possible direction for deeper analysis

CSC JumpStart Service is provided into two halves

First part is **Data Fitness**

Second part is Analysis (not discussed here)

- Assessment of data structure and adequacy of the m5 reviewer guide
  - Standard domain consistency (e.g., is the AE outcome in a separate domain?)
  - Custom domains and their content (e.g., are derived efficacy endpoints in SUPP domains?)
  - Adequate description of issues in the m5 reviewer guide

- Identify data issues that could impact review
  - Can I use standard review tools (e.g., JReview, MAED)?
  - Can I run common analyses (e.g., liver function, Hy's Law plot)?
  - What other data quality issues should I be aware of that could impact review?

- What our tools can tell us and our eyes confirm
  - Are appropriate variables available?
  - Where is the missing data?
  - Where are all the deaths (if any)?
  - Did the sponsor use standard terminology?
  - Is data well described by metadata?
  - Coding assessments by term matching algorithms?
  - Unit matching and standardization (SI or conventional)?

4

We put data fitness issues into four categories/objectives to help with filing review

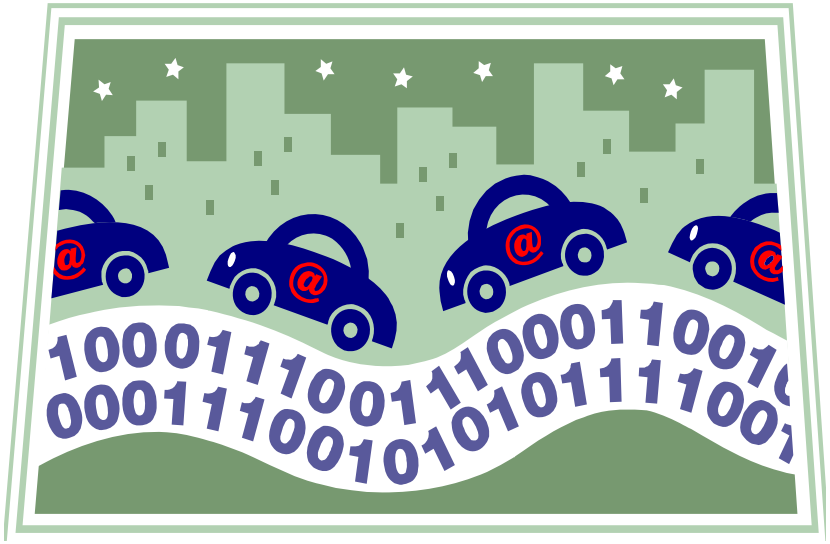
1. **Navigation and composition** – determine where the safety and efficacy data elements exist in the standard and custom domains and how well this is documented in the Reviewer's Guide (SDRG) in m5 (e.g., What domains contain death information)

2. **Data Integrity and Completeness** – determine if data is reliable, consistent, complete (e.g., Are all patients accounted for? Is there missing data?)

3. **Reviewability** – determine if available tools can readily be used or whether you will need additional programming support to perform review

4. **Data Traceability** – determine where or whether the data in the CRF is represented (ex. Is data for deaths, serious events and adverse dropouts consistent with CRF data?)

We chose these categories to assist in the completion of the Clinical



DATA FITNESS Objective 1:
NAVIGATION/ COMPOSITION

Center for Drug Evaluation and Research
CSC
Computational Science Center

STANDARD DOMAINS
Data in the same place, w/ the same name & meaning – parent domains (where are deaths?)

CUSTOMIZED & SUPPLEMENTAL
linked to parent domains (derived variables, efficacy, unique to TA)
Described in **SDRG** (ADRG*)

TA	Trial Arms	Trial Design
TE	Trial Elements	Trial Design
TI	Trial Inclusion/Exclusion Criteria	Trial Design
TS	Trial Summary	Trial Design
TV	Trial Visits	Trial Design
CO	Comments	Special Purpose
DM	Demographics	Special Purpose
CM	Concomitant Medications	Interventions
EX	Exposure	Interventions
SU	Substance Use	Interventions

ZC	Right Heart Catheterization	Findings
ZI	Independent Adjudication	Findings
ZW	Walk Test	Findings
RELREC	Related Records	Special Purpose
SUPPAE	Supplemental Qualifiers (AE)	Special Purpose
SUPPCE	Supplemental Qualifiers (CE)	Special Purpose
SUPPCM	Supplemental Qualifiers (CM)	Special Purpose
SUPPDM	Supplemental Qualifiers (DM)	Special Purpose
SUPPDS	Supplemental Qualifiers (DS)	Special Purpose

*Coming soon

Finding	Impact
Deaths and Disposition: a. Information about deaths can be recorded in **AE, DM, DS**, custom, or supplemental domains.	Death counts may not align with CSR or ISS data.
a. Disposition events like lack of efficacy or withdrawal to an adverse event have often been coded to 'Other'.	Analysis panels can be skewed, and signals diluted by inappropriate coding.

Example:

Not in **DS**

DOMAIN	USUBJID	DSTERM	DSDECOD	EPOCH	DSSTDTC	DSSTDY
DS	A	ADVERSE EVENT	ADVERSE EVENT	MAINTENANCE	date1	200
DS	B	COMPLETED	COMPLETED	POST-TREATMENT	date2	259

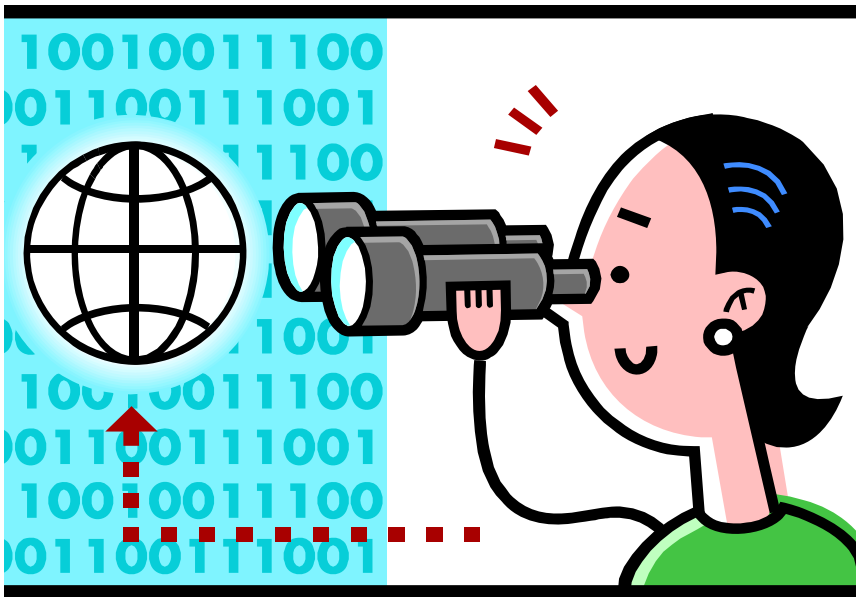
Some Detail in **AE**

DOMAIN	USUBJID	AETERM	AEDECOD	AESER	AEREL	AEOUT	EPOCH	AESTDTC	AESTDY
AE	C	Unknown cause of death	Death	Y	NO	FATAL	POST-TREATMENT	date3	280

Custom Domain **XX**

DOMAIN	USUBJID	XXSEQ	XXSTDTC	XXTERM	XXCAT	XXSCAT	XXEVAL
XX	D	1	date4	Death			
XX	D	11	date4	Unknown	CCV	MACE	ADJUDICATION COMM

Finding	Impact
Unique Domain - **SUPPYY** Contains radiographic reports and tumor lesion size at baseline	This will support the efficacy endpoint
Unique Domain – **RELREC** Does not adequately link between radiographic reports and tumor size domains **TUMSZ** **LDIAM** Longest Diameter **LPERP** Longest Perpendicular.	Unable to replicate change in lesion size



DATAFITNESS Objective 2:
DATA INTEGRITY AND COMPLETENESS

Data Quality

Data Attributes:

- Wrong date format, missing data,
- Terminology, nonstandard codes
- Consistent use of units

Data INTEGRITY:

- Completeness
- Screen failures – availability in I/E
- Is data content consistent when represented across domains ? (Disposition Data, Dropouts, Deaths)

12

Finding	Impact
Mismatched or missing terms	This will require additional review time to investigate

Example

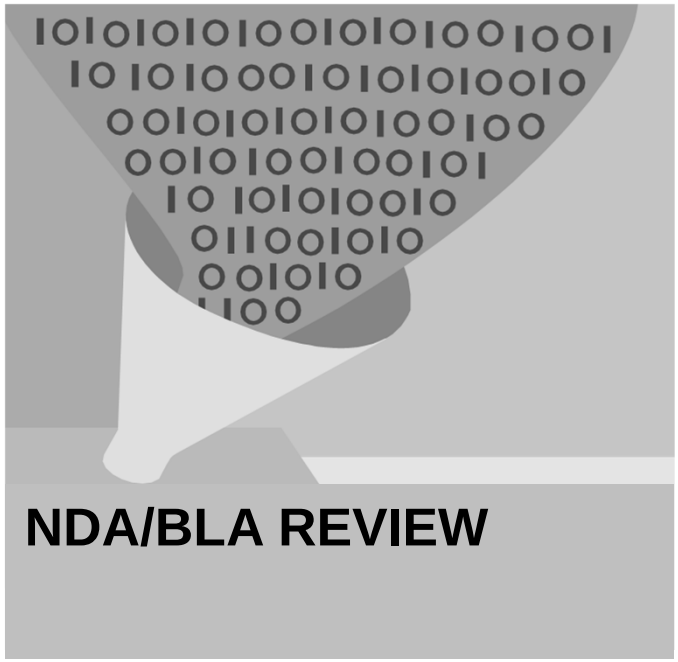
Study A

ID	AETERM	AEDECOD	AESER
17	STROKE		
18	STROKE		
19	Left clavicular communited (*sic*) fructure (*sic*) with displacement		Y
20	MYOCARDIAL INFARCTION		
21	MYOCARDIAL INFARCTION		
22	STROKE		

Study B

ID	AEDECOD	PT_NAME
23	Functional gastrointestinal disorder	Functional residual capacity

Finding	Impact
Missing data: a. Patient X has data in **CO, DS, custom domains, DV, SE, SUPPDS, SUPPDV, SV)** but not in **DM** a. Patients Y and Z have missing **EX** data but have data in **AE** b. Patient W has 8 **EG** records at VISITNUM=199, but is not in **DM**.	Reviewer needs to determine if these patients were in the Safety Analysis Population (Explanations of these issues would be ideal for the Reviewer Guide in m5 (SDRG)



DATAFITNESS Objective 3
REVIEWABILITY

Center for Drug Evaluation and Research
Computational Science Center

Can more readily assess reviewability for SAFETY

- Analytic needs common across NDAs
- Tools for safety analyses developed and used
- Standard terminology & SDTM representation

Reviewability for EFFICACY currently limited to description of component data

- Analyses customized for specific therapeutic area
- Need additional tool requirements (R, SAS etc)
- Need robust endpoints, data standards, uniform protocols

Finding	Impact
AGE and DOB is not provided for 33% and 20.7% of subjects in Study A and B.	Limits analysis by age subset.
Missing Baseline flags for labs: • ALT - 96.6% • ALP - 97.1% • AST - 97.1% • BILI - 95.4%	Required for tool use, unable to perform liver assessments, Hy's law outputs

DATA FITNESS Objective 4
DATA TRACEABILITY

FIT DATA …

- Is traceable to the source (EHR = CRF =source)
- Derived data is supported by SDTM components
- Analyses from the SDTM data approximates results from ADaM data and tables in CSR

Finding	Impact
Traceability of SAE: **SAE** information not collected in CRF. **SUPPAE** refers to "*Date SAE report submitted to sponsor*" Study: # records w/o CRF trace • Study A = 129 • Study B = 127	Reviewer is unable to trace the Serious Adverse Event data to the CRF
Screen Failures do not have **IE** records.	Reviewer is unable to detect bias in patient selection

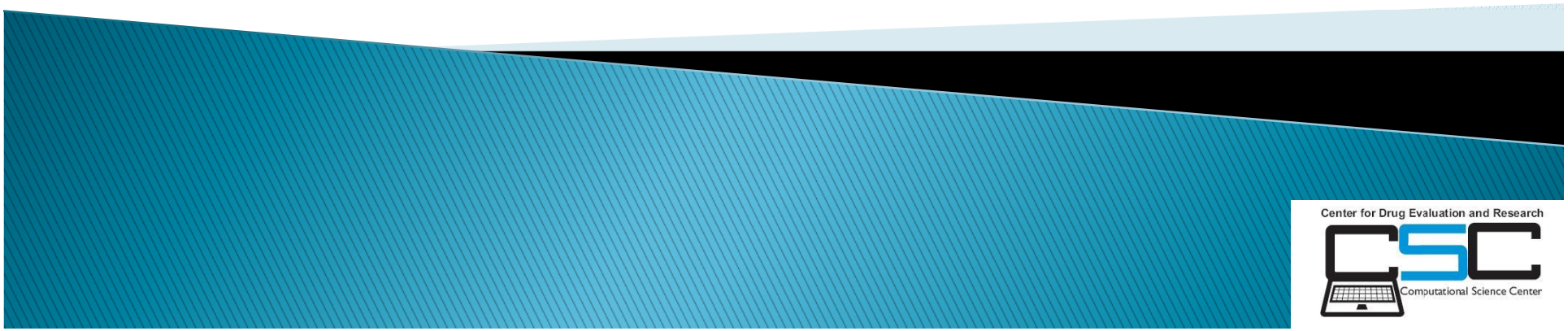
In Conclusion

Higher percentage of data is coming through the gateway with SDTM data at CDER

Industry is getting better at implementing the standard

Standardized data helps CSC help the Review Team

Questions?


Center for Drug Evaluation and Research
CSC
Computational Science Center