An Overview of the Office of Computational Science Core DataFitness Service

Tejas Patel[1]; David Jacobs[2]; Qingying Lu[2]; Kathryn Matto[2]; Jamal Horne[2]; Janie Ma[1]
[1]FDA CDER OTS Office of Computational Science, [2]IBM

What is the OCS Core DataFitness service?

The Office of Computational Science (OCS) provides CDER reviewers innovative and reliable solutions that improve and strengthen the scientific review process by integrating data, tools and training. This poster details OCS's Core DataFitness (CoreDF) service, designed to help reviewers understand the quality of clinical study data, including identifying data conformance issues to data standards, early in the review cycle. The CoreDF service includes:

- A set of data quality reports consisting of an overview tab and issue-specific reports
- A deskside support session to help reviewers navigate the reports

Summary Section

The summary section contains useful links to important documents, at-a-glance study information such as MedDRA version, subject counts and metadata about the data package.

Reports Section

Reports are provided to expedite common medical reviewer tasks. These include reports related to subject deaths, adverse event and disposition coding quality reports, and a report listing supplemental domains and variables included with the submission. These reports enable medical reviewers to complete common review tasks more quickly and without having to organize, sort or join data on their own. These reports also highlight potential records of interest that would otherwise be very time-consuming to identify.

Adverse Events Coding Quality Report

The CoreDF reports includes an adverse event coding quality that contains a comparison of the Reported Terms (AETERM) against the Lower Level Terms (AELLT) and Dictionary-Derived Terms (AEDECOD). An algorithm generates a "score" from 0 ("Could not match") to 100 ("Direct Match"). Review teams can use this report to:

- Check how a specific reported term is coded
- Check all terms coded to a specific AEDECOD
- Systematically review all mapping

Application: NDA123456
Study: ABC-001

Back to Summary

Adverse Events Coding Quality

Description
This report will help you methodically select a sample of adverse events to examine coding quality. The algorithm uses approximate string matching and does not have medical background. The score represents similarity between Reported Term (AETERM) and either MedDRA PT (AEDECOD) or MedDRA LLT (AELLT) in the submitted data. A score of 100 means the strings are identical, while a score of 0 means that algorithm was unable to determine sufficient similarity. Many terms that have 0 or low scores will, in fact, be coded properly. Using this report will allow you to cut down on the number of terms to review by cutting out those with higher scores.

MedDRA 19.1
MedDRA version used for this report was pulled from define.xml. Report could show additional mismatches if there is a discrepancy between the MedDRA version in define.xml and the actual MedDRA version used to code adverse events.

Reported Term (AETERM)	MedDRA LLT (AELLT)	MedDRA PT (AEDECOD)	Match Details	Score	Number of Rows
Right Side Abdominal Pain	Abdominal Pain Localized	Abdominal pain	Could not match	0	1
Subcostal Pain	Right Upper Quadrant pain	Abdominal pain upper	Could not match	0	1
Worsening of Anemia	Anemia Aggravated	Anemia	Partial word match to PT	35	2
Creatinin Increase related to the treatment	Creatinine increased	Blood creatinine increased	Partial word match to PT	35	1
Mandibula pain	Bone pain	Bone pain	Partial word match to PT	35	1
Intermittent Runny Nose	Runny nose	Rhinorrhoea	Partial word match to LLT	60	1
Intermittent oral mucositis	Mucositis oral	Stomatitis	Partial word match to LLT	60	1
Blood bilirubin increased	Blood bilirubin increased	Blood bilirubin increased	Direct match to PT	100	3
Fatigue	Fatigue	Fatigue	Direct match to PT	100	68

How does CoreDF relate to other OCS services?

CoreDF is an important component in the OCS service portfolio which includes JumpStart, KickStart, the OCS Service Desk and others. Currently, CoreDF serves as the primary effort to provide automated data quality reporting to FDA medical reviewers for all new applications. It is also a complementary service to the JumpStart and KickStart services, both of which provide high-touch, in-depth support on data quality and safety analyses tailored to specific applications for a limited number of applications. OCS ensures that its portfolio of services meets the wide variety of reviewer needs through both automated processes and high-touch, tailored approaches.

Overview Tab of the Core DataFitness Reports

Application: NDA123456
Study: ABC-001

A Phase III, Randomized, Double Blind, Placebo Controlled Study of New Therapy X

Summary

Documents
Study Data Reviewer's Guide
Define.xml

Standards / Dictionaries
SDTM-IG 3.1.3
SDTM-CT 2017-09-29
MedDRA 19.1

Subjects / Actual Arms
30 - Subjects
15 - Treatment mg/kg/day (50.0%)
15 - Placebo mg/kg/day (50.0%)

Datasets
42 - Total Datasets
3 - Custom Datasets
14 - Suppqual Datasets

Findings

Demographics
- 5 (16.7%) of randomized subjects were not treated
- 2 (6.7%) of randomized subjects are missing Subject Reference End Date/Time (RFENDTC)
- 1 (< 0.1%) of subjects are missing Date/Time of Informed Consent (RFICDTC)

Disposition
- 8 (0.9%) of disposition statuses or protocol milestones are potential duplicates

Exposure
- 14 (2.0%) of treatments occurred after Date/Time of Last Study Treatment (RFXENDTC)

Adverse Events
- 6 (< 0.1%) of adverse events have neither severity or toxicity grade populated
- 70 (12.0%) of events are missing end time-point

Laboratory
- 1 (< 0.1%) of baseline observations are missing Standard Results (LBSTRESC)

Vital Signs
- No significant findings

Other
- 22 (100.0%) of derived variables in Define.xml are missing computational algorithms
- EPOCH variable was not provided

Reports To Help Basic Review Activities

Deaths
- Death Summary
- Death Details
- Death Reconciliation

Adverse Events
- Adverse Events Coding Quality

Disposition
- Disposition Coding Quality

Supplemental Info
- Supplemental Contents

Findings Section

The findings section is organized by domain – Demographics, Disposition, Exposure, Adverse Events, Laboratory, Vital Signs and Other. Each finding listed in the 'Findings' section is hyperlinked to a detailed report of that finding.

Other Core DataFitness Efforts

Partnerships: OCS seeks to partner with more offices within CDER who are interested in automated data validation in order to reduce the time it takes to understand the submission data package. OCS has already partnered with the Office of Clinical Pharmacology and the Office of Oncological Diseases to meet the more specific needs of these offices.
SDTM to ADaM Traceability: OCS is currently working to include traceability reporting as part of the CoreDF service. This reporting would help reviewers trace Analysis Data Model (ADaM) data elements back to their SDTM origins.
Nonclinical Data: OCS will expand the CoreDF service to support nonclinical reviewers by providing a set of reports identifying data conformance issues to Standard for Exchange of Nonclinical Data (SEND) in nonclinical studies early in the review cycle.

Deskside Support

CoreDF analysts meet with medical reviewers to guide them through all aspects of the CoreDF reports. This includes explanations of all summary, report and findings sections. Analysts also guide medical reviewers through each finding in detail, explaining the meaning or potential impact of the finding. CoreDF analysts are also able to address medical reviewer's questions in the session. These meetings and the opportunities to guide reviewers through their CoreDF reports therein constitute the service component of the CoreDF service.

Example Finding – Not Treated Subjects

This example finding pertains to subjects who have values for ARM and ACTARM that indicate they were treated but who are missing records in the Exposure (EX) domain. This could indicate that the subjects were not treated or that the subjects are truly missing exposure information. Not treated subjects should have null values for ACTARM, and all subjects populated with a treatment value for ACTARM should have records in the EX domain.

No Exposure record found for subject

Finding
5 (16.7%) of randomized subjects were not treated

Impact
Missing useful information to speed up your own exploration and analysis.

USUBJID	ARM	ACTARM
000-000-001	OBSERVATION COHORT	OBSERVATION COHORT
000-000-005	OBSERVATION COHORT	OBSERVATION COHORT
000-000-007	OBSERVATION COHORT	OBSERVATION COHORT
000-000-011	OBSERVATION COHORT	OBSERVATION COHORT
000-000-015	OBSERVATION COHORT	OBSERVATION COHORT

Example Finding – Missing AE Severity/Toxicity

This example finding pertains to records in the Adverse Events (AE) domain missing values for Severity (AESEV) or Toxicity (AETOXGR). Severity or toxicity is a required variable and should be collected on the CRF and transmitted to the dataset. Records missing severity/toxicity may be excluded from certain analyses or lead to discrepancies in results.

Neither AESEV or AETOXGR is populated

Finding
6 (< 0.1%) of adverse events have neither severity or toxicity grade populated

Impact
Missing values lack analytical value.

USUBJID	AESEQ	AETERM	AEDECOD	AETOXGR	AESTDTC
000-000-009	2	axillary vein thrombosis	Axillary vein thrombosis		2015-04-10
000-000-009	3	herpes labialis	Oral herpes		2015-04-04
000-000-027	3	Cystitis	Cystitis		2015-06-22
000-000-027	4	Cystitis	Cystitis		2015-07-06
000-000-030	5	Herpes of nose-upper mouth	Herpes virus infection		2015-03-23
000-000-030	7	Pain to the joints of the hands	Arthralgia		2015-10-14

Conclusion - Benefits of CoreDF Service

The CoreDF service helps reviewers understand the quality of clinical study data, including identifying issues related to conformance to data standards, early in the review cycle. By leveraging the CoreDF service reviewers increase their understanding of the overall SDTM data package, leading to more informed communications with the applicant and the identification of potential information requests (IRs).

Office of Computational Science (OCS)

Better Data. Better Tools. Better Decisions

For More Information Please Contact:
Janie Ma, MS, PMP
Program Manager, CoreDF
Office of Computational Science
Center for Drug Evaluation and Research
U.S. Food and Drug Administration
email: janie.ma@fda.hhs.gov