**PhUSE US Connect 2018**

**Paper RG13**

# The Changing Landscape of Data Quality at CDER: an Overview of the Core DataFitness Assessment

Crystal Allard, FDA, Silver Spring, MD, USA
Michelle Dieter, FDA, Silver Spring, MD, USA
Lilliam Rosario, FDA, Silver Spring, MD, USA

## ABSTRACT

The Core DataFitness Assessment (CoreDF) is the first step toward providing data quality information to all CDER reviewers early in the review process. CoreDF reports are generated automatically and provided to review teams without a JumpStart Session. They are intended to be standalone and self-explanatory. The CoreDF reports provide approximately 70% of the data quality information previously provided only through a JumpStart Service. This paper describes the CoreDF reports and some of the feedback we've received from reviewers who have received them.

## INTRODUCTION

The CoreDF Assessment is comprised of a set of reports that provide automated data quality information. CoreDF Assessments are a new way to convey data quality information for more CDER applications in a shorter timeframe and with fewer resources than the JumpStart Service can provide.They are intended to be standalone and self-explanatory. The Office of Computational Science (OCS) is piloting the CoreDF Assessment in FY 2018 and plan to roll them out fully in FY 2019. Ultimately they will be automatically provided for all applications with SDTM data.

## BACKGROUND

FDA/CDER has attempted to enforce receipt of study data that can be received, archived, and analyzed in various ways, including data standards, utilizing data models to enforce quality of data within a warehouse, refusal to file, and technical rejection. Study data standards have been utilized for this purpose for more than 10 years at FDA. Enforcing conformance to these standards is a step toward ensuring data quality. However, enforcing conformance to a standard can never fully ensure that data can be understood, analyzed, and assessed for fileability, safety, and efficacy. Though CDER has been able to run a validator on SDTM data for years, the reports have not previously been appropriate for CDER reviewers. Previous validation reports were large and difficult to comprehend, especially for medical officers who don't have a background in data science. Previous iterations of the FDA validator foused heavily on assessing conformance to a standard. The advent of conformance rules provided by SDOs and the business rules written by FDA have started a journey toward stratifying conformance rules and review-relevant findings.

OCS provides services to help clinical reviewers in the Regulatory Review process. The JumpStart Data Fitness Service has been successfully providing information about the quality of standardized study data to about 30 applications per year for the last 5 years. Using the JumpStart Service, clinical reviewers can better understand the data within an application to

conduct an effective evaluation of the drug submission. The Data Fitness portion of the JumpStart Service helps clinical reviewers assess the “reviewability” of clinical trials study data by evaluating data composition and quality, and determining the types of analyses thatcan be performed and which analytical tools can be used.

Providing JumpStart Services is resource-intense. The JumpStart Data Fitness Session is one of 2 JumpStart Sessions and comprises about half of the information presented during a JumpStart Service, but utilizes signicanlty more than half of the resources it takes to provide the full service. Preparing for a JumpStart DataFitness Session takes 2 weeks and includes multiple analysts, managers, and tools. OCS has been urged by reviewers and leadership in the Office to New Drugs, CDER to provide the service to many more reviewers. Due to the success of the JumpStart Serivce, OCS has explored new ways to convey data quality information for more CDER applications in a shorter timeframe and with fewer resources than the JumpStart Service can provide.

The Core Data Fitness Assessment includes a set of core data quality findings that have been idenitified by the JumpStart team as common. It focuses not on conformance to a standard, but on review-relevant findings that will help a reviewer quickly identify areas for further exploration. The CoreDF Assessment helps reviewers understand whether the data for a particular study is complete and analysis-ready. They also help reviewers use their time exploring quality findings to determine wether a finding is truly an issue and whether that issue will required further action, such as conveying it to a sponsor in an information request.

## HOW IT WORKS/HIGH LEVEL PROCESS

The Core DF reports were created utilizing a list of common data quality errors found in JumpStart to reduce the full list of SDTM validation rules and business rules to a digestible report of data quality issues for a reviewer. In FY 2018 The Core DF reports are provided to reviewers receiving a JumpStart Service in addition to those who request a service but are not able to receive it. Agile methodologies are being used to gather feedback, develop and enhance the reports, and plan for expansion in FY 2019.

### Overview of DataFit Core Service Delivery

The DataFit Core Service utilizes the FDA’s customized DataFit validator to produce the CoreDF Assessment. In FY18, this service is provided to every reviewer that receives a JumpStart Service. The reports are provided prior to the JumpStart Service and reviewed during the JumpStart Data Fitness Session. For reviewers who are not able to receive a JumpStart Service, a CoreDF Assessment is provided separately with an hour of desk-side support to help the reviewer navigate and understanding the reports. While offering desk-side support, the CoreDF team hears first-hand whether a particular reviewer is able to understand and use the CoreDF reports. This one on one feedback is particularly helpful to the development team in understanding and empathizing with the needs of reviewers. This helps the team assess whether the reports can be provided to a reviewer in the absence of the support provided during a JumpStart Service.

## Design Principles

While developing the Core Data Fitness Assessment, the team followed a set of design principles to guide decision-making. These principles include:

1) Fully Automatable – All CoreDF Reports can be run automitically with a tool without human intervention

2) Review Relevance – CoreDF Reports will not focus on conformance to a standard, but rather on findings that will provide information of value to a reviewer

3) Reviewer Focused Language - The CoreDF reports need to use language that a reviewer will understand without a list of acronyms or dictionary. Ideally, a reviewer will be abel to utilize the CoreDF reports without receiving training on any specific data standard

4) Actionable - CoreDF Findings will provide sufficient information that a reviewer can understand how to explore a finding and determine next steps

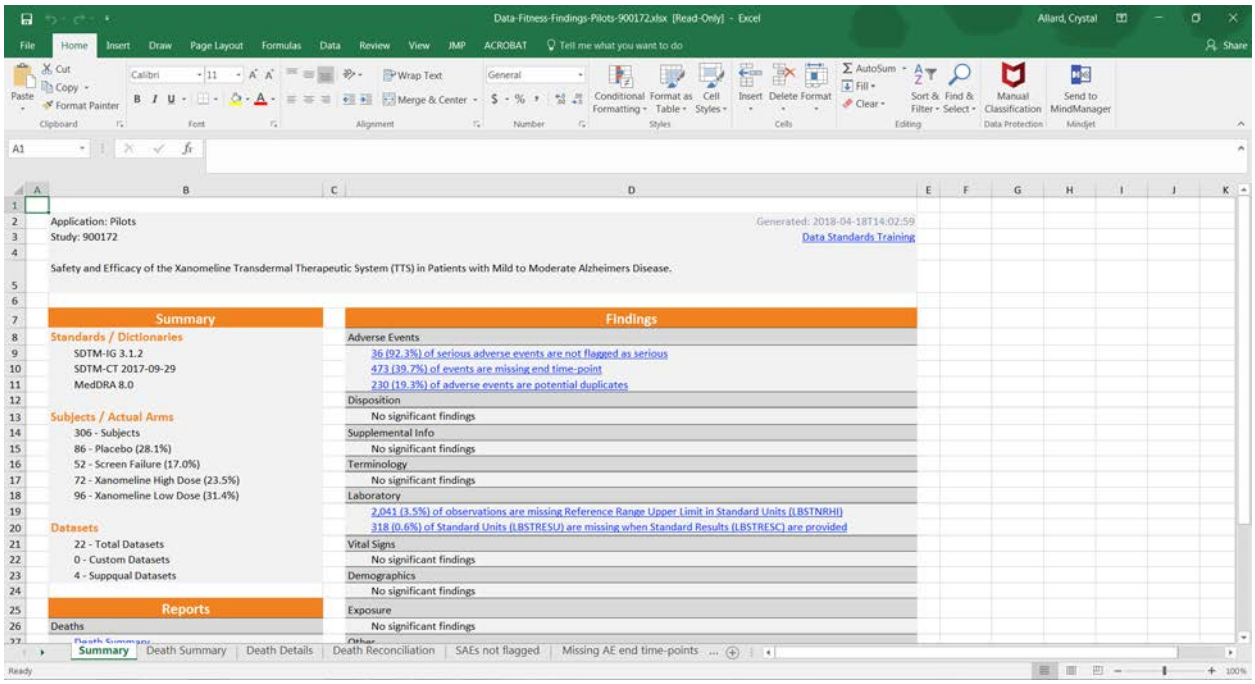
5) Exploratory – CoreDF reports provide findings such that reviewers have the information they need to explore relevance. They do not prescribe relevance nor tell a reviewer what they ought to think about a particular finding.

6) Truth without Interpretation – CoreDF reports present truthful findings without interpretation.

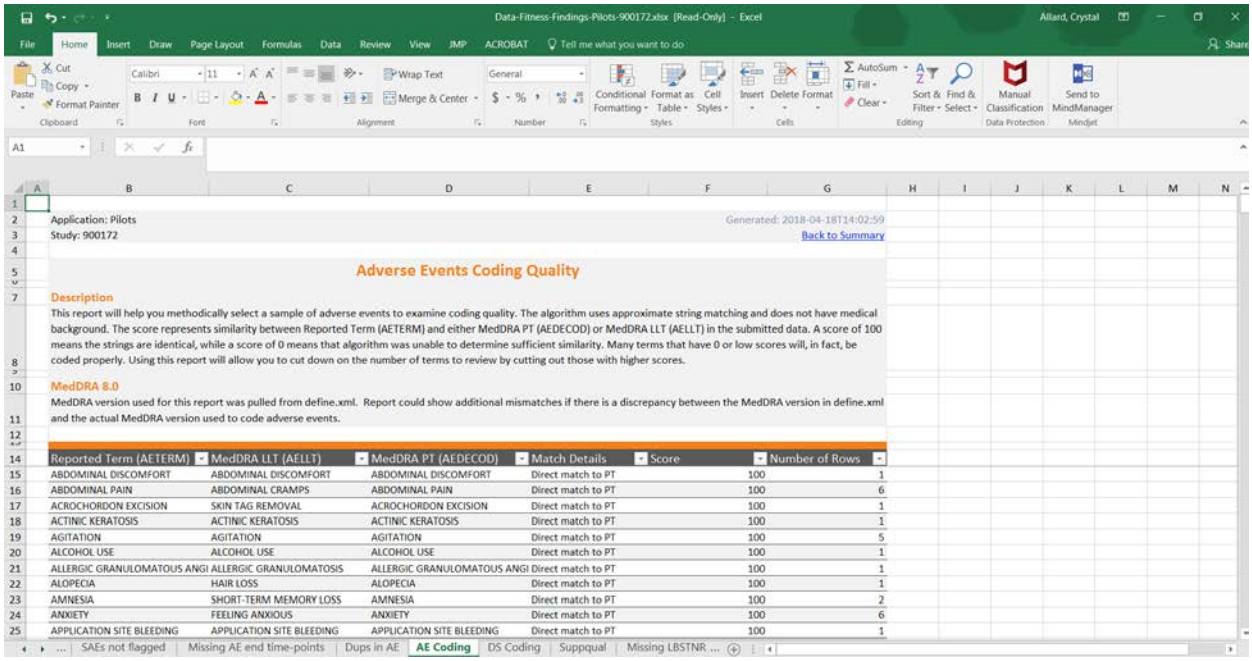
## Description Of The Reports

The Core Data Fitness Assessment is an excel-based output from the OCS Validator Tool, DataFit. It consists of a summary page and 19 separate reports.

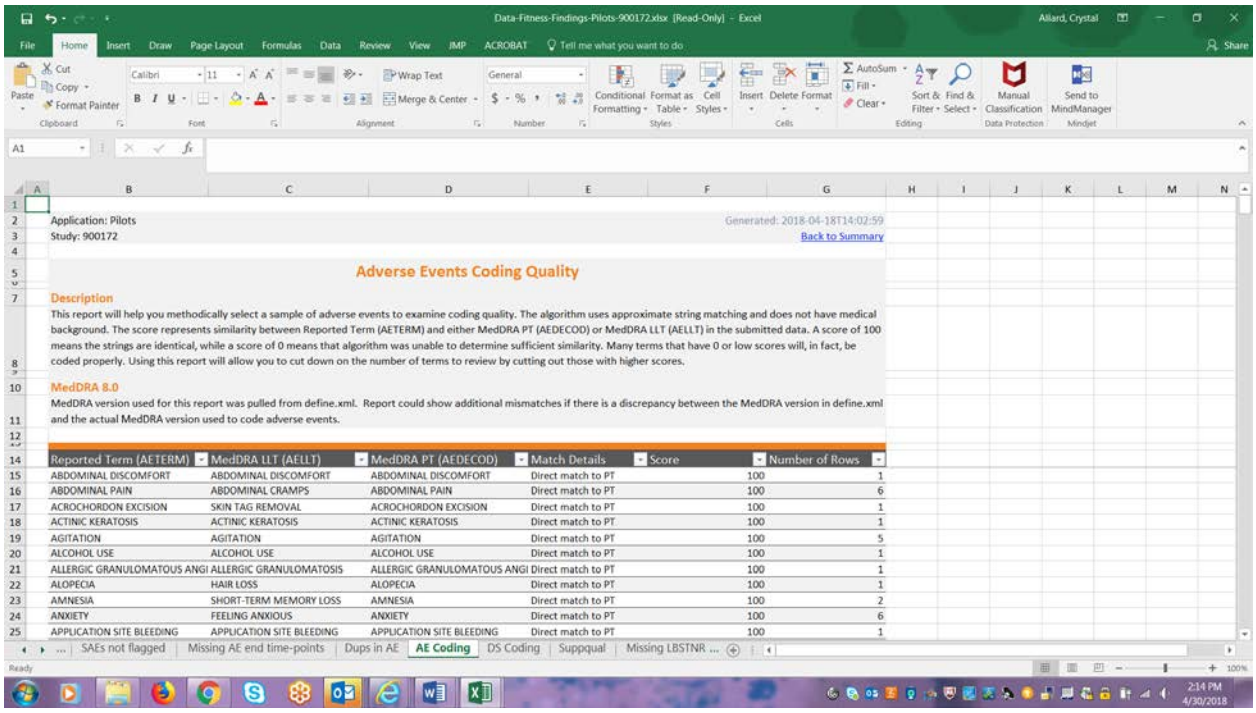
The summary page includes information about the versions of standards and terminologies used, the number of subjects and arms, and basic information about the datasets. It also includes high level information about the various reports included and the findings within each report.




Each tab contains a specific report with information about data quality findings. There are reports for death information, AE Coding, duplicates, supplemental qualifiers information, and much more. An example of the AE coding report is below:




The supplemental qualifiers report describes the information contained within supplemental datasets:



Application: Pilots
Study: 900172
Generated: 2018-04-18T14:02:59
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 8.0

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 |
|---|---|---|---|---|---|
| ABDOMINAL DISCOMFORT | ABDOMINAL DISCOMFORT | ABDOMINAL DISCOMFORT | Direct match to PT | 100 | 1 |
| ABDOMINAL PAIN | ABDOMINAL CRAMPS | ABDOMINAL PAIN | Direct match to PT | 100 | 6 |
| ACROCHORDON EXCISION | SKIN TAG REMOVAL | ACROCHORDON EXCISION | Direct match to PT | 100 | 1 |
| ACTINIC KERATOSIS | ACTINIC KERATOSIS | ACTINIC KERATOSIS | Direct match to PT | 100 | 1 |
| AGITATION | AGITATION | AGITATION | Direct match to PT | 100 | 5 |
| ALCOHOL USE | ALCOHOL USE | ALCOHOL USE | Direct match to PT | 100 | 1 |
| ALLERGIC GRANULOMATOUS ANGI | ALLERGIC GRANULOMATOSIS | ALLERGIC GRANULOMATOUS ANGI | Direct match to PT | 100 | 1 |
| ALOPECIA | HAIR LOSS | ALOPECIA | Direct match to PT | 100 | 1 |
| AMNESIA | SHORT-TERM MEMORY LOSS | AMNESIA | Direct match to PT | 100 | 2 |
| ANXIETY | FEELING ANXIOUS | ANXIETY | Direct match to PT | 100 | 6 |
| APPLICATION SITE BLEEDING | APPLICATION SITE BLEEDING | APPLICATION SITE BLEEDING | Direct match to PT | 100 | 1 |

SAEs not flagged | Missing AE end time-points | Dups in AE | AE Coding | DS Coding | Suppqual | Missing LBSTNR ...

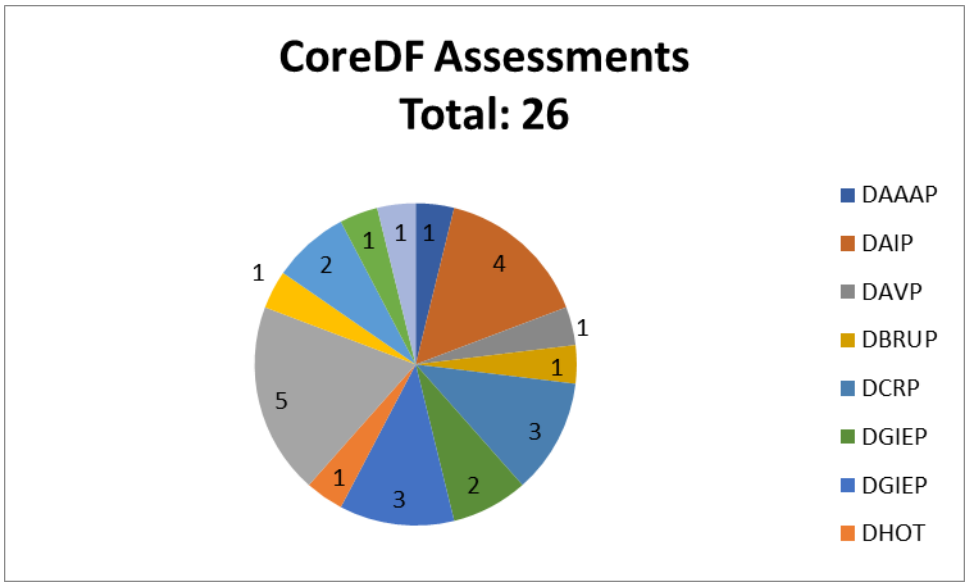
## Benefits To Reviewers

On average, 63% of Data Fitness findings that JumpStart is presenting to reviewers are being shown in Core Data Fitness Assessments.

Reviewers get a set of data quality assessments that will help them identify data quality issues early in their review. Because the reports are automated, they can be generated very early in the review process. These reports are intended to be self-explanatory so a review can digest them without the aid of an analyst.

## Analysis Of Feedback

Core Data Fitness Assessments were provided to 25 JumpStart applications in between November 2017 and April 2018.




Following receipt of the Core Data Fitness Assessment, each reviewer is asked to provide feedback on the ease of use of the reports. The reports are currently provided prior to the jumpstart service which is dependant on availability of resources but generally occurs prior to filing. 85% of medical reviewers found the information within reports relevant for that time point in their review. In the future, the CoreDF Assessment will be provided within the first two weeks of receipt of an application containing study data.

The CoreDF assessment is intended to be standalone and self-explanatory. 100% of reviewers said information within reports was easy to navigate as well as understand. The CoreDF reports are currently provided to reviewers either as part of a JumpStart Service or with support from a Data Fitness analyst, so it's hard to tell whether the feedback reflects reports that are truly standalone. As the reports are provided to a larger audience in the future, without support from a JumpStart analyst, it will be important to continue to gather and analyze this feedback. 90% of reviewers felt nothing was missing from reports. One reviewer ask for a link to the "define file either directly in the reports or a hyperlink to it". As the CoreDF Assessment matures, better navigation functionality will be added.

## CONCLUSION

The Core Data Fitness Assessment was created to give reviewers an automated overview of the quality of their study data prior to the filing meeting. By adhereing to our design principles, following agile development cycles, and assiduously gathering and reviewing user feedback, we have developed a set of reports that are ready to be rolled out to reviewers on a large scale in fiscal yearl 2019.

The CoreDF Assessment will help reviewers understand and explore the quality of their study data early in the review process. This will allow them to make decisions abtou the fileability of the data in an application and whether they need to send information requests to a sponsor in order to analyze and review the data in order to make a decision regarding safety and efficacy of the product.

## REFERENCES

Assessment of Impact of Electronic Submissions and Data Standards on the Efficiency and Other Performance Attributes of the Human Drug Review Process

Data Standards Strategy 2015-2017 - FDA

**FDA Resources for Data Standards**

FDA Office of Computational Science

## RECOMMENDED READING

Common Data Quality Issues Identified During the JumpStart Process

Common Errors in Loading SDTM Data to the Clinical Trials Repository - Why Getting it Right Matters

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Crystal Allard
FDA
10903 New Hampshire Ave.
Silver Spring, MD 20903
Work Phone: 301-796-8856
Email: crystal.allard@fda.hhs.gov
Web: www.fda.gov

*This paper reflects the views of the author and should not be construed to represent FDA's views or policies.