PhUSE US Connect 2019

Paper SA03

High Quality Study Data Standards for Submission

Majdoub Haloui, Merck & CO., Inc, North Wales, PA, USA
Ellen L. Asam, Merck & CO., Inc, North Wales, PA, USA

ABSTRACT

High quality study data standard deliverables are essential to support the regulatory review process. Delivering CDISC compliant submissions require understanding the regulatory agencies requirements and CDISC standards. Deliverables are complex and require compliance to the individual CDISC standards as well as consistency and agreement between multiple components. While industry models and templates provide structure and guidance and validation tools help to identify some quality issues, however they are not comprehensive and many quality issues can be overlooked.

This paper intends to help sponsors prepare quality, CDISC compliant submissions. The paper focuses on providing detailed steps for a quality review for data and documents that can be followed to ensure regulatory and CDSIC compliance.

INTRODUCTION

A key factor to support review and approval for any submission is preparing high-quality study data standard deliverables free of issues or inconsistencies. Issues or inconsistencies may result from a lack of knowledge of the latest agencies requirements, non-compliance to the CDISC standards and inconsistencies between different documents in the submission package. This paper will provide a high-level overview of regulatory requirements and components of CDISC submission data packages. And based on a thorough review of current regulatory requirements and our experience, we will propose detailed steps for a quality review for data and documents. While this material is generally applicable to any study data standard package for submission, the paper will focus primarily on submission of clinical data.

REGULATORY AGENCIES REQUIREMENTS

FDA

FDA has instituted new requirements for data standards that will apply to most study data submitted to FDA's Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER).
Beginning after the dates specified below, FDA may refuse to file for New Drug Applications (NDAs) and Biologics License Applications (BLAs) or refuse to receive for Abbreviated NDAs (ANDAs) any electronic submission whose study data do not conform to the required standards specified in the FDA Data Standards Catalog.
In this section we will outline key FDA requirements for regulatory submissions.

1. **FDA Guidance**
   - *Providing Regulatory Submissions in Electronic Format - Standardized Study Data: Guidance for Industry (PDF - 136KB) (Dec. 2014)*
   - *Providing Regulatory Submissions in Electronic Format - Submissions Under Section 745A(a) of the FD&C Act: Guidance for Industry (PDF - 81KB) (Dec. 2014)* –
2. **FDA Data Standards Catalog**
   The *FDA Data Standards Catalog* contains specifications for the data standards and versions that the FDA supports. The supported data standard versions are listed along with dates support begins and ends.
3. **Technical Rejection Criteria for study data**
   Technical Rejection Criteria for Study Data
4. **FDA Technical Conformance Guides**
   - *Study Data Technical Conformance Guide v. 4.2*
     This guide provides technical recommendations to sponsors for the submission of animal and human study data.

- *Vaccines Technical Specification Guidance v1.0*
  This document provides detailed information and specifications for the content of datasets submitted to FDA's CBER Office of Vaccines Research and Review (OVRR).

5. **FDA Business Rules/FDA Validator Rules**
   The *Business Rules* help ensure that the study data are compliant, useful, and will support meaningful review and analysis.
   The *Validator Rules* are used by the FDA to ensure data are standards compliant and support meaningful review and analysis.

PMDA

PMDA is currently in a transitional period. During this transition period, PMDA accepts legacy submissions and partial data submissions (hybrid submissions). Sponsors need to have consultation meetings (consultation on data format of submission of electronic study data) with PMDA one year before the submission. During that meeting, Sponsors need to gain agreement with PMDA about the electronic datasets for NDA submission.
After this transition period, starting April 1, 2020, all required study data will need to be submitted in CDISC format. No waivers will be allowed after this date. All clinical study data meeting eStudy submission criteria must be compliant with CDISC standard format for submissions. Therefore, closed or completed studies in legacy format will need to be converted.

The Office of *Advanced Evaluation with Electronic Data* website page provides required items for submission study data to the PMDA.

1. *Data Standards Catalog*
   Data Standards Catalog contains a list of acceptable versions of Data Exchange Standards and Terminology Standards that PMDA supports.
2. *Technical Conformance Guide on Electronic Study Data Submissions*
   The Technical Conformance Guide provides technical details for electronic study data submission.
3. *Basic Principles on Electronic Submission of Study Data for New Drug Applications*
   Document and Q&A about basic principles for electronic submissions of study data to PMDA.
4. *Notification on Practical Operations of Electronic Study Data Submissions*
   Additional information about the electronic submission of study data to PMDA
5. *Study Data Validation Rules*
   PMDA has published a set of Study Data Validation Rules for SDTM, ADaM, and define.xml. Each of these rules is assigned a severity level Reject, Error or Warning.
6. *FAQs on Electronic Study Data Submission (Excerpt)*
   The FAQs on Electronic Study Data Submission summarizes inquiries on electronic study data submission received by the PMDA in a Q&A format.

OTHER REGULATORY AGENCIES

1. **European Medicines Agency (EMA)**
   The EMA does not require individual patient data to be submitted electronically.
2. **China Food and Drug Administration (cFDA)**
   cFDA has recently started to accept the submission of electronic data to support drug applications but does not yet mandate the adoption and adherence to industry wide data standards or submission of data in the eCTD backbone.
3. **Health Canada (Canada)**
   Canada does not require submission of electronic data but have asked for data for selected submissions.
4. **Ministry of Food and Drug Safety (Korea)**
   Korea has asked for data and programs for select submissions.

COMPONENTS OF CDISC SUBMISSION DATA PACKAGES

1. STUDY DATA STANDARDIZATION PLAN

The Study Data Standardization Plan (SDSP) is a document intended to communicate to CDER and CBER the expected CDISC (SDTM, ADaM and SEND) standards used for data in upcoming clinical and pre-clinical trials.

In addition, for clinical studies that will be submitted to CBER, the SDSP CBER appendix should be provided to the review office no later than the End-of-Phase 2 meeting. The CBER SDSP appendix should outline proposed SDTM domain/variable usage, supplemental domain usage and proposed analysis.

2. SDTM DATA PACKAGE

The key components of an SDTM Submission package are the SDTM annotated case report form (acrf.pdf), datasets in xpt format, the data definitions document (define.xml v2.0) and the Study Data Reviewer's Guide.

a) aCRF

aCRF is a document that maps the clinical data collection fields used to capture subject data (electronic or

paper) to the corresponding variables contained within the SDTM datasets. It should be provided as a PDF with the file name 'acrf.pdf'. aCRF is linked through the define.xml when the variable origin is set to 'CRF'. aCRF should be submitted preferably at the time a protocol is submitted.

b) Datasets

The SAS Transport Format (XPORT) Version 5 is the file format required for the submission of all electronic datasets.

c) Define.xml

Define.xml is an XML document that provides the metadata for human and animal model datasets using the SDTM and/or SEND standards. It is based on CDISC Define-XML Specification. The define.xml is the most important part of the electronic dataset submission for regulatory review. All SDTM data must be accompanied with a define.xml. The define.xml displays the submitted datasets metadata in a user-friendly format that can be programmatically reviewed for conformance and consistency with the study data.

d) Study Data reviewer's guide (SDRG)

The SDRG should facilitate a reviewer's use of the submitted data and help the reviewer understand the relationships between the study report and the data. It should describe any special considerations for the study, such as explanations for any unusual data points, or unique circumstances (ex. decisions and variations resulting from migrated or external data or a legacy data conversion). It should also include explanations for any data conformance issues. The study data reviewer guide for clinical studies should be named csdrg.pdf and nsdrg.pdf for nonclinical studies.

3. ADAM DATA PACKAGE

The key components of an ADaM Submission package are the datasets in xpt format, the data definitions document (define.xml v2.0), the Analysis Study Data Reviewer's Guide (adrg.pdf) and software programs. Analysis Results Metadata (analysis-results-metadata.pdf) and Supplemental Data Definitions Document if needed.

a) Datasets

The SAS Transport Format (XPORT) Version 5 is the file format required for the submission of all electronic datasets. The XPORT is an open file format published by the SAS Institute for the exchange of study data. There should be one dataset per transport file.

b) Define.xml

Define.xml is an XML document that provides the metadata for datasets using the ADaM standards. It is based on *CDISC Define-XML Specification*. The define.xml is the most important part of the electronic dataset submission for regulatory review. All ADaM data must be accompanied with a define.xml file. The define.xml displays the submitted datasets metadata in a user-friendly format that can be programmatically reviewed for conformance and consistency with the study data.

c) Analysis Data Reviewer's Guide (ADRG)

The preparation of an ADRG is recommended as an important part of a standards compliant analysis data submission for clinical trials. It purposefully duplicates limited information found in other submission documents, to provide reviewers with a single point of orientation to the analysis datasets. The ADRG provides reviewers with context for analysis datasets and terminology, beyond what is presented in the define.xml. Like the SDRG, it should describe any special considerations for the study, such as explanations for any unusual data points, or unique circumstances. The ADRG also provides a summary and explanations of ADaM conformance findings. The ADRG should help a reviewer to understand the data and to reproduce primary analyses, datasets and major endpoints. The ADRG should be named 'adrg.pdf'.

d) Software programs

Sponsors should provide the software programs used to create all ADaM datasets and tables and figures associated with primary and secondary efficacy analyses. The specific software utilized should be specified in the ADRG. Sponsors should submit software programs in ASCII text format. Executable file extensions should not be used. Software programs should include sufficient documentation to allow the reviewer to understand the submitted programs.

e) Analysis Results Metadata

Analysis Results Metadata (ARM) is a document that identifies the critical analyses and provides traceability from results in a statistical display to the analysis data and programs used to create it. It documents the analyses performed and provides the regulatory reviewers with the means to trace results back to their source documents (programs, datasets, and statistical analysis plans). PMDA is already asking for an ARM (TCG section 4.1.2.1 Definition documents of datasets). While the ARM is not a mandatory requirement for FDA many reviewers find it useful.

f) Supplemental Data Definitions Document (This is an optional document)

Define.xml does not support complex formatting. If a derivation description is complex, requires formatting (tables, equations, multiple paragraphs) or is too long to be displayed in the comment/derivation cell of the define.xml then it should be documented in an external document that is linked to the variable in the define.xml. Currently, there is no standard naming convention for this external document, but some sponsors have been using 'complex-algorithms.pdf' for the file name.

DETAILED STEPS FOR A QUALITY REVIEW FOR CDISC SUBMISSION DATA PACKAGES

High quality study data and accompanying metadata are extremely important for the regulatory review process. They add significant value to the submission by allowing reviewers to efficiently interpret and understand submitted data. The result is a time savings for the regulatory reviewer, which can potentially expedite the drug approval process. The following recommended checks should be performed to ensure quality. These detailed steps have been compiled based on a careful review of current industry requirements.

1. ACRF

aCRF is a main reference for reviewers to understand how data was collected. Perform a thorough QC of annotated pages compared to actual data. It is recommended to think about automating this process as only automated tools can catch inconsistencies when comparing define.xml or data against aCRFs.

The validation checks described below should be performed. Ensure the following:

- Regardless of whether the clinical database is legacy or SDTM compliant, an aCRF should be submitted
- No spelling errors in dataset labels and variable names
- The annotated CRF should include all available forms used to collect source data but only unique CRF forms should be annotated
- Annotation of SDTM variables, not raw data variables
- aCRF should not be a scanned document and should be searchable
- File name should be 'acrf.pdf' not 'blankcrf.pdf'
- The text "NOT SUBMITTED" should be annotated on the case report form where data are recorded on the CRF but not submitted, and an explanation for why data was not submitted should explained in the Study Data Reviewer's Guide
- All collected variables in the domain datasets (xpt files) should be accounted for in the aCRF (No missing annotations)
- If there is more than one domain per page, each set of domain variables should be annotated in its own color
- Supplemental domains should be the same color as the parent domain
- Every domain on the aCRF page should have the 2-letter domain code at the top left of the page, followed by the domain decode that matches the dataset label
- All text in the annotations that represent variable and domain names should be capitalized
- If possible, the annotations should not obstruct any text on the CRF page.

2. DATASETS

The study data should conform to CDISC SDTM, ADaM, SEND standards and sponsor's custom standards. The data should be validated using the latest version of Pinnacle 21 (P21) software used by the agency. Additionally, the validation checks described below should be performed. Ensure the following:

a) Datasets should be created using PROC COPY with the XPORT engine and not PROC CPORT/CIMPORT

b) Label and length restrictions:
   - All datasets and dataset variables should have labels
   - Dataset labels should be unique across SDTM & ADaM
   - For all datasets, the variable length should be set to the maximum length of the variable used
   - Dataset labels should clearly describe the content of the dataset
   - All dataset and variables should have length according to the following table:

Dataset label/variable	Maximum Length in Characters
Dataset Label	40
Variable Name	8
Variable Label	40

c) Variable names should contain only uppercase letters, numbers, and must start with a letter. Dataset names should contain only lowercase letters, numbers, and must start with a letter. No other symbols or special characters should be included in these names. Variable names, as well as dataset and variable labels should include American Standard Code for Information Interchange (ASCII) text codes only

d) Dataset and variable labels can include punctuation characters. However, special characters should not be provided, such as:
   - Unbalanced apostrophe
   - Unbalanced single and double quotation marks
   - Unbalanced parentheses, braces or brackets
   - '<' less-than sign and '>' greater-than sign

e) All required and expected variables should be present in the dataset

f) All Permissible variables for which data were collected (or necessary for derivations) should be present in the dataset
g) All Permissible variables that have NULL or missing values should be dropped from the dataset unless they were intended to be collected on the CRF
h) Dataset Size:
    - Datasets greater than 5 gigabytes (GB) in size should be split into smaller datasets no larger than 5 GB in size
    - If you need to split a file that exceeds file size limits, you should submit the smaller split files in the 'split' sub-folder in addition to the larger non-split file in the original data folder
    - Datasets that are split should be clearly named ex: LB1, LB2 and LB3
    - There is no need for a second define.xml file to be submitted within the split subfolder
    - The explanation regarding how these datasets were split should be documented in the define.xml/ reviewers guide
i) File names should only contain letters, numbers or hyphens no underscores or spaces
j) All file names should be in lower case

3. DEFINE.XML

The structure and content of the define.xml for both tabulations and analysis should be checked for accuracy, clarity and traceability. The define.xml should be validated using the latest version of the Pinnacle21 tool used by the agency. Additionally, the validation checks described below should be performed. Ensure the following:

a) Define.xml format

- The define.xml should be created as per CDISC Define-XML Specification Version 2.0 and should not be compressed
- Conformance to CDISC/sponsor standards should be adhered to in terms of dataset structure, dataset/variable/value-level metadata and controlled terminology
- A style sheet should be included and placed in the same submission folder as the define.xml file

b) Datasets order, Class, Structure, Keys and Location

- Datasets should be listed in alphabetical order by name attribute within each class in the define.xml file. The recommended order for the classes is:

SDTM	ADaM
1. Trial Design Datasets 2. Special-purpose Domains 3. Interventions Domains 4. Events Domains 5. Findings Domains 6. Findings About 7. Relationship Datasets	1. ADSL 2. Basic Data Structure 3. ADaM Other

- All datasets submitted are listed in the define.xml and vice versa
- All dataset attributes (description, structure, class, keys etc.) should be populated in the defne.xml
- Class attribute should have a value of 'SPECIAL PURPOSE', 'FINDINGS', 'EVENTS', 'INTERVENTIONS', 'TRIAL DESIGN', or 'RELATIONSHIP' for SDTM and SEND data and 'SUBJECT LEVEL ANALYSIS DATASET', 'BASIC DATA STRUCTURE', or 'ADAM OTHER' for ADaM data.
- Structure should match the mapping specification for SDTM datasets and ADaM specification for ADaM datasets. Structure text should align with values listed in the key column
- The dataset keys listed in the define.xml should be "natural keys" and they should identify a unique record within a dataset. If natural keys identify more than one record within a dataset, then data and keys should be carefully reviewed to check whether it is a result of duplicates within the data or incorrect keys. The sequence of variables in the key should be consistent with the structure of the dataset
- The link in the location column should point to the dataset location

c) Variable order, Label, Data Type, Controlled Terms, Origin, Derivation and Comment

- Variable order should match with the dataset (xpt file) variable order
- The variable length in the define.xml should match with the variable length in the dataset
- Variable origin type attribute should only have a value of 'CRF', 'Derived', 'Assigned', 'Protocol', 'eDT', or 'Predecessor'
- If a variable has multiple origins/datatypes, leave the origin/datatype blank at the variable level and provide origin/datatype at the value level metadata
- If the origin of a variable is 'CRF', CRF page numbers must be populated for that variable.
- If the origin of a variable is 'Predecessor', predecessor must be provided
- If the origin of a variable is 'Derived', derivation must be populated for that variable
- The explanation of derivations in the "Source/Derivation/Comment" column (or in computational algorithm section) should be written in English instead of programming code. The derivations should be easily understood by any reviewer
- Derivations should be clear and concise to help replicate the results

- Derivations and comments should not be blindly copied from the mapping/ADaM spec into the define.xml
- Derivations and comments should not reference raw data or local files, the reviewers do not have access to your raw database or local files
- The controlled terminology should be provided in the define.xml for all variables that have controlled terminology associated with it as per the SDTM-IG/ADaM-IG version being used for the study. Additionally, sponsor-defined controlled terminology should be provided in the define.xml
- Controlled terms (CDISC and Sponsor defined) must be specified in the codelist section of the define.xml. In the rendered define the controlled terms column links to the table of terms. If the controlled terminology is based on an external dictionary (e.g., MedDRA, WHODRUG), a link to the external dictionaries section of the define should be provided
- UNIT codelist should be unique for each unique unit variable in the study
- Codelist should list all planned terms for the variable – all CRF options should be included in the codelists, not just those present in the data

d) Bookmarks and hyperlinks

- The define.xml should have bookmarks for navigating to external documents such as "Annotated Case Report Form", "Study Data Reviewer's Guide" for SDTM define and "Analysis Data Reviewer's Guide", "Analysis Results Metadata" for ADaM define
- Bookmarks and hyperlinks should be checked to verify that they are functioning correctly, are appropriately labeled/identified, and point to the correct locations
- The styles sheets associated with the define.xml should be provided along with the define.xml

4. DATASET AND DEFINE VALIDATION

Study data validation helps to ensure that the study data are compliant, useful, and will support meaningful review and analysis. Study data should be validated against CDISC standards, eCTD and FDA/PMDA validation rules. In addition, the validation checks described below should be performed:

- Ensure compliance with Technical Rejection Criteria for both FDA and PMDA
- Define.xml version 1.0 is no longer supported and define.xml version 2.0 should be used for any submission after March 15, 2018
- Validation of the datasets and define.xml must be performed using the latest Pinnacle21 version used by the agency. The P21 validation should be performed two different ways. One way is to perform validation on the data and define and another way is to perform validation on just the define.xml file. The results of the validations should be reviewed carefully to ensure that there are no rejects, errors or warnings related to issues in the data, the define.xml or both
- The define.xml file should open without any issues in the internet browser and all bookmarks and links in the define.xml should navigate correctly to their destinations
- All dataset attributes (description, structure, class, keys etc.) should be populated in the defne.xml
- The description of datasets in the define.xml should exactly match with internal label of datasets. Similarly, variable label presented in the define.xml should exactly match with the variable label in the dataset
- The keys listed for datasets should represent a unique record within a dataset
- The variable order in the define.xml should match with the variable order in the dataset
- The variable length in the define.xml should match with the variable length in the dataset
- All data values of value level metadata variable (e.g. LBTESTCD) should be present in the define.xml
- The define.xml should be checked to make sure that there are no orphan controlled terms, computational algorithms, or value level metadata variables (e.g. controlled term section has a codelist that is never referenced for any variables in the define.xml)
- FDA/PMDA validation issues severity

Item	FDA	PMDA
Rejects	All rejects must be resolved before submission.	Rules which, if violated, will cause the review to be suspended until corrections have been made. All rejects must be resolved before submission.
Errors	• Errors that **CAN** be fixed **SHOULD** be fixed • Errors that cannot be fixed should be documented and explained in the reviewer's guide with a clear rationale	Rules which, if violated without any prior explanation, will cause the review to be suspended until corrections have been made.
Warnings	• Warnings that **CAN** be fixed **SHOULD** be fixed • Warning that cannot be fixed should be documented and explained in the reviewer's guide with a clear rationale	Rules which, even when violated, will not necessarily require any explanation.

P21 Enterprise or community version is the leading application used by sponsors and CROs to validate SDTM /

ADaM datasets and define.xml against CDISC standards. Both FDA and PMDA are using Pinnacle 21 Enterprise to review submission data from sponsors. P21 tool does not cover all the checks listed above, so additional quality checks software need to be created to supplement the tool's assessments to ensure accuracy, validity and completeness of SDTM/ADaM datasets as well as the define.xml. Adding manual quality checks can be cumbersome and time consuming.

5. REVIEWERS GUIDE

This document describes any special considerations or directions to quickly orient the reviewers to the submitted data and help them understand the relationship between the study report and the SDTM or ADaM data. The reviewers guide should be well formatted, polished, and professional. It should be concise and clear, easy to read, use and navigate. It is important to ensure the information provides sufficient detail without being overly wordy or confusing. The following quality checks should be performed:

a) Reviewers guide template from PhUSE

Follow the recommended the template from PhUSE and fill out all sections. Do not delete any sections that do not apply to your study, use Not applicable (N/A) instead.

b) Format

Check for the following and ensure that none of these formatting issues are present in the reviewer's guide

- Inconsistent fonts, size, color
- Missing or incorrectly working hyperlinks
- Different formats used across tables
- Orphan records or headers when a table continues across a page
- Hidden or special characters copied from other documents
- Inconsistent line spacing or indentation
- The ADRG is not placed with the analysis data in Module 5 of the eCTD or the file name is not correct.

c) Content

The content of SDRG and ADRG should be carefully checked for accuracy and clarity.

- All sections should be populated correctly
- All questions should be answered correctly and match the data
  - Data cutoff date, if provided, should match the cutoff date in the TS domain
  - Screen failure domains in the SDRG should match with the screen failure domains in the submitted data
  - List of domains planned, but not submitted should match the annotated CRF
- List of key variables from the SUPPxx domains that have contributed partly or fully to the generation of ADaM datasets should be provided
- Ensure all information provided is relevant. Do not include extraneous information such as internal information (names of people or references to other studies not part of the submission)

d) Writing Quality

- Ensure correct spelling, grammar and punctuation
- Use a consistent verb tense
- Use a consistent voice and person

e) Data Conformance Issues

Any data conformance issues should be explained in the Data Conformance Summary section. This helps reviewers understand how to use the data during analysis and preempts the need for additional questions and clarifications.

- Fix all issues that can be fixed and provide meaningful explanations for data conformance issues that remain
- Ensure the explanation includes detail of why any data issue remains
- Always use the latest version of Pinnacle 21 software used by the agency

f) Cross Reference checks between SDTM define, SDRG, ADAM define and ADRG

Check for general consistency of information between documents:

- Protocol number, title and design information should be consistent between SDRG and ADRG
- The information provided in SDRG section 1.2 (Study Data Standards and Dictionary Inventory), ADRG section 1.3 (Study Data Standards and Dictionary Inventory) and define.xml should be consistent
- The information provided in the SDRG section 3.3 (SDTM Subject Domains) should be consistent with SDTM define.xml
- The information provided in the ADRG section 5.3 (Analysis Datasets) should be consistent with ADaM define.xml

6. ANALYSIS RESULTS METADA (ARM)

The ARM assists reviewers by identifying critical analyses and supports understanding and reproducing analysis results. Therefore, the structure and content of the ARM should be checked for accuracy and clarity. Below is a list of ARM checks that should be performed:

- All tables/figures listed in ARM are present in the CSR

- All programs referenced in ARM are present in the programs folder and linking correctly
- All dataset and variables referenced in ARM are present in the dataset folder and linking correctly

7. ELECTRONIC SUBMISSION FORMAT - ECTD DIRECTORY STRUCTURE

- Study datasets and their supportive files should be organized into a specific file directory structure when submitted in the eCTD format (See Figure 1 and Table 1)
- Submission of files within the appropriate folders allows automated systems to detect and prepare datasets for review, and minimizes the need for manual processing
- If you need to split a file that exceeds file size limits, you should submit the smaller split files in the 'split' sub-folder in addition to the larger non-split file in the original data folder
- Do not submit empty file folders and additional subfolders
- If additional folders are needed, consult with the appropriate center in advance for guidance
- Refer to TCG for a detailed Study Dataset and File Folder Structure and Description

DISCUSSION

Submission of study data standards involves many components and has become a complex process. It is a significant challenge to ensure all the pieces are accurate, complete and consistent. At Merck we are working to continuously improve our process.

We have implemented additional automated quality checks beyond what is available in the Pinnacle 21 Enterprise tool, to ensure consistency between the define.xml documents and the submission data. These automated supplemental checks are far more efficient than performing manual review. For more information please refer to the eSubmission - Are you really Compliant? Paper that was presented in PharmaSUG 2018.

We have also instituted a review process for the Data Reviewers Guides consisting of multiple levels of review. First, authors and team members, who are familiar with the study data and analysis, review documents to ensure accuracy and completeness of content. Then we have a review by members of management, including individuals who are less familiar with the study details, to ensure clarity, the level of detail is appropriate, and that the documentation overall makes sense. While it can be difficult to complete this activity in a timely fashion, we have found and addressed many issues in the documents which have improved quality.

CONCLUSION

High quality study data and accompanying metadata are extremely important for the regulatory review process. They add significant value to the submission by allowing reviewers to efficiently interpret and understand submitted data. The result is a time savings for the regulatory reviewer, which can potentially expedite the drug approval. The list of quality checks presented in this paper can help expedite the quality review time.

REFERENCES

Office of Advanced Evaluation with Electronic Data (PharmaSUG 2018 April 29 - May 2, Seattle, Washington)
https://www.pmda.go.jp/english/review-services/reviews/advanced-efforts/0002.html

Electronic Data Submission and Utilization in Japan
https://www.lexjansen.com/pharmasug/2018/REG/PharmaSUG-2018-REG03-PPT.pdf

FDA Bioresearch Monitoring Technical Conformance Guide
https://www.fda.gov/downloads/drugs/developmentapprovalprocess/formssubmissionrequirements/ucm332468.pdf

Standardized Format for Electronic Submission of NDA and BLA Content for the Planning of Bioresearch Monitoring (BIMO) Inspections for CDER Submissions
https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequire%20ments/ucm332466.pdf

FDA Technical Rejection Criteria for Study Data
https://www.fda.gov/downloads/drugs/developmentapprovalprocess/formssubmissionrequirements/electronicsubmissions/ucm523539.pdf

PhUSE-DataSubmissions-FAQ
https://phuse.teamworkpm.net/#/files/2106425

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Please contact the authors at:

Majdoub Haloui
Merck & Co., Inc
351 N. Summneytown Pike
North Wales, PA 19454
Work Phone: 267-305-2629
Email: Majdoub.haloui@merck.com
Web: www.merck.com

Ellen Asam
Merck & Co., Inc
351 N. Summneytown Pike
North Wales, PA 19454
Work Phone: 267-305-6803
Email: ellen_asam@merck.com
Web: www.merck.com