PharmaSUG 2020 Paper SS-156

Analysis Package e-Submission – Planning and Execution

Abhilash Chimbirithy, Merck & Co. Inc., Rahway, USA
Saigovind Chenna, Merck & Co. Inc., Upper Gwynedd, USA
Majdoub Haloui, Merck & Co. Inc., Upper Gwynedd, USA

ABSTRACT

Analysis package is one of the e-Submission components submitted to regulatory agencies as part of INDs, NDAs, ANDAs, and BLAs. Analysis package contains the study analysis data and related files following a standardized electronic format. Proper planning and resources are required to create this package which has several individual components such as datasets in XPT format, define, analysis results metadata (ARM), data reviewers guide, and programs. These components are separate deliverables but interrelated. With frequently identified challenges, questions and issues from previous studies, we have provided guidance related to planning, setting deliverables timelines, identifying team members' responsibilities and ensuring regulatory required data standards compliance for the components submitted. In this paper, we will present details for best practices, proper planning and checklists that can help teams efficiently create an analysis package. It also highlights ways to achieve effective cross-functional collaboration and consistently meet regulatory compliance.

INTRODUCTION

This paper discusses analysis package which is one of the main components of CDISC based e-submissions. It highlights different components required within analysis package, how these different components can be planned early, streamlined and created successfully to achieve a compliant regulatory submission. Having worked on multiple analysis package and successful regulatory submissions, we have listed challenges encountered during the process and solutions to address them. This paper highlights Involvement of cross functional teams and how effectively to improve collaboration, tools that are helpful to create the components effectively in timely fashion, stakeholders involved and responsibilities. It also provides examples to further help programmers and statisticians involved in analysis package e-submission process.

ADAM SUBMISSION PACKAGE COMPONENTS

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

- **ADaM Datasets**
  The SAS® Transport Format (XPORT) Version 5 is the file format required for the submission of all electronic ADaM 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.

- **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. Stylesheet (xsl) file is accompanied with xml document so define.xml can be viewed properly.

- **Analysis Data Reviewer's Guide (ADRG)**
  The preparation of an ADRG is recommended as an important part of a standard's 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 Clinical Study Data Reviewer's Guide (cSDRG), 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'.

- **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 enough documentation to allow the reviewer to understand the submitted programs.

- **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). ARM is not a mandatory requirement for FDA submissions, but many reviewers find it very useful. However, PMDA strongly recommends submitting an ARM document for all its submissions.

- **Supplemental Data Definitions Document (optional)**
  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 derivation 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.

CROSS-FUNCTIONAL TEAMS AND RESPONSIBILITIES

Statisticians and Statistical Programmers
Collaboration between these two teams are critical to complete the deliverables efficiently and with high quality. Both teams are involved from the early stages of the study analysis and reporting activities. Creating ADaM dataset specifications for the study is a collaborative effort and all team members must ensure description of analysis variable derivations are accurately written to support define.xml creation. The specifications are authored by either the statistician or statistical programmer, depending on the type of dataset. Typically, statistician is the author for efficacy analysis datasets and special safety event analysis datasets specifications, and algorithms for key variables in other datasets, if needed. Attempt must be made to write derivation description that could be easily understood by a non-programmer. A well written ADaM dataset specifications at the initial stages of the study saves a lot of time during ADaM datasets development and analysis package creation. If company specific standard templates exist, then it must be followed. Except for special safety event analysis datasets, the statistical programmer is the author for non-efficacy analysis datasets. Datasets specifications need to be thoroughly reviewed by both programmer and statistician. Any changes to the dataset specifications should be captured in the revision history section and ensure to alert relevant stake holders.

Clinical and Statisticians
Statisticians lead the efforts to consult with clinical study team members and regulatory liaison to finalize the list of primary and key secondary end points for the study, for which related programs and analysis results metadata will be created by the programming team.

Data Management and Statistical Programmers
Since ADaM Datasets are derived from SDTM datasets, issues that persist in SDTM are inherited to ADaM datasets. Pinnacle 21 validation checks identify issues related to ADaM datasets. These issues could be due to collected data, issues with mapping of collected data to SDTM or ADaM datasets not following the ADaM principles. It is important for statistical programming team to obtain inputs related to data issues from data management. Any issues that cannot be fixed at a data level must be explained in the issue summary of conformance section of ADRG.

Regulatory Liaison and Regulatory Publishing
It is important to obtain the submission deliverables timelines from the company regulatory liaison. Programming team can use the timelines to determine the resources required to complete the package. Regulatory publishing team is responsible for assembling all the submission packages before sending it through electronic common technical document (eCTD) interface to agency. To ensure that submission packages are successfully sent through the

gateway, submission packages must follow the eCTD guidelines, such as folder structure, file naming conventions and file locations. Teams must be aligned and in agreement with publishing team regarding expectations for the analysis package.

Submission related Subject Matter Expert (SME) and Study Teams

Submission SMEs champion and ensure submission readiness and quality. They share study data standards (CRT SDTM package / ADaM Analysis package) knowledge with the study team, stay current on complex and rapidly evolving regulatory study data standards submission requirements and establish & improve the e-Submission process. It is critical for the study team to continuously consult and obtain advise on the most current regulatory standards from submission SMEs. Additionally, the study team should engage the submission SMEs to conduct quality review of study data standards deliverables (SDTM Tabulation and Analysis packages).

CHALLENGES

1. Specifications or Metadata

There are many challenges encountered during the iterations of drafting ADaM datasets specification and generating analysis datasets. Following is a list of common issues related to the specifications:

- Inconsistencies between data and metadata
- Spelling mistakes
- Incorrect or lack of derivation details
- Derivations rules not in plain language leading to inconsistent interpretation across functional groups
- Not using latest templates created by the company
- Use of special characters in the derivation description
- Creating parameter level metadata
- Not performing thorough validation or review by the validation programmer or reviewer (statistician)

2. Datasets

At the time of submissions, several issues are seen in ADaM datasets. Team members not understanding the ADaM principles leads to creation of non-compliant ADaM datasets. Few team members choose to perform P21 validation closer towards the data base lock or after the data base lock, identifying numerous issues at the end. At this time, it may be too late to fix all issues. There are few checks that are not implemented within P21, but issues exist in ADaM and must be fixed before submission to regulatory agencies. Attributes for SDTM variables are updated when retaining them in ADaM datasets, leading to consistency and traceability issues between SDTM and ADaM. Inconsistencies are also identified within ADaM datasets, such as same variable having different attributes. Few teams create a huge ADSL dataset with codes running multiple pages, leading to complexities and challenges at the end.

3. Tool (Pinnacle21) issues

- Usage and training
- Adapting to newer versions
- Interpretation of issues
- False positive issues

4. Coordination between multiple teams and proper communication

- Setting up effective communication with data management related to data issues identified when creating ADaM datasets
- Challenges explaining the ADaM principles to other functional groups (statisticians, Clinical, and Data Management)
- Addressing Pinnacle 21 identified issues before database lock with inputs from Clinical, Statistician and Data Management, document issues that cannot be resolved

5. Planning and Timeline

- Coordinating multiple stake holders review and addressing their comments
- Establishing review cycle and completion timelines

6. Supplemental Data Definitions

- When and what to include

7. ADRG – what is too much or too little

- Derivation details which are not covered in specifications (define)
- Content Quality
- Review Quality
- Compliance Section Details

8. Analysis Results Metadata (ARM)

- Identifying key analysis results that needs to be included
- Collaborating between statistician and programmer
- Heavy document with metadata information for all TLFs included

SOLUTIONS

1. Specifications or Metadata

Automatizing the dataset attributes directly from the ADaM dataset specifications would reduce a lot of downstream problems, issues and errors. Variables labels, lengths and formats can easily be applied directly to ADaM datasets from the specifications by developing applicable macros. Based on feedback and suggestions from programmers, ADaM dataset specifications have constantly evolved to incorporate automatic checks and logics within excel workbook to make sure variables name and label don't exceed 8 and 40 characters respectively. Conditional checks were introduced in ADaM dataset specification template columns such as type and origin, so user cannot enter the values other than the ones expected. Ex: Type column can only have Char, Text, Integer and Num. Similarly, Origin column can have values only Predecessor, Derived and Assigned.

One must accurately identify key variables for each dataset when creating the specification. The key variable identified can subset unique records from the dataset. Typically core variables identified in ADSL must be carried forwarded to all other datasets with consistent attributes, should be automatized to avoid inconsistency.

2. Datasets

Standard operating procedure and template programs when applicable should be in place to ensure all team members responsible for creating ADaM datasets complete trainings related to ADaM principles and guidance. It's very important for study teams creating ADaM datasets follow ADaM IG principles and check for compliance with Pinnacle 21 tool as soon as draft ADaM datasets are ready. A second subject level optional dataset ADBASE can be created to hold baseline characteristics related variables that is an addition to what is present in the ADSL. It can include baseline characteristic variables that can be derived/copied from other ADaM datasets. This step can help ADSL to be less complex

Special tools and utilities can be developed to check some of the common errors and inconsistencies such as attributes consistency between specifications and data, consistency among predecessor variables, finding special characters in the specifications.

3. Tools (Pinnacle21)

Developing adequate trainings and work instructions for Pinnacle 21 end-to-end, will enable efficient use of the tool and achieve consistent ADaM compliance. When a user is provided access to Pinnacle 21, roles and responsibilities must be clearly defined and understood. Identifying Subject Matter Experts (SME) facilitates answering additional questions related to the tool and any issue in achieving ADaM compliance. Documenting known Pinnacle 21 compliance issues related to ADaM data and supplying standard text to be used in reviewers guide for the issues not resolved would greatly enhance users' productivity and efficiency when finalizing ADRG submission package.

4. Coordination between multiple teams, establishing consistency and efficient communication

Coordination between statistician and programmer is essential when ADaM dataset specifications are finalized and identify clear roles and responsibilities between statistician and programmer with respect to safety or efficacy datasets specifications. It's essential for programmer to understand SAP and mockup TFL shells. It is important to get clarification from statistician in case of any questions related to the above-mentioned documents which in turn drives designing appropriate and correct ADaM datasets. Study programming lead ensures to run Pinnacle 21 validation checks as soon as draft ADaM datasets are available. Identify issues and triage accordingly based on

category of issues. Execution resources and SMEs help establish consistency and efficient communication

- Execution Resources (ER)

Execution Resources specific to ADaM implementation can be created to achieve consistency across study teams. Below is a list that we have developed in our organization and this has proven to be very helpful to team members creating ADaM and related analysis packages in a consistent manner.

ADaM Kick-of slides – Orientation providing End to End process and resources available to support ADaM implementation.

Guidelines document for ADaM dataset specifications.

ADaM Checklist for e-Submission

ADaM Guidelines for Integrated Analysis.

ADaM guidelines for Cross over studies.

Pinnacle 21 Tool related ERs – Work instructions specifying the usage of tool, ADaM data validation and define.xml creation instructions.

- SME

It is good to establish a team of SMEs. They can provide consultation to their respective colleagues on standards, Pinnacle 21, and submission deliverables. SMEs serve as a point of contact for any questions, issues, feedback, suggestions for improvement. Monthly SME meetings help leads to highlight any gaps, feedbacks, suggestions from their respective teams. At our organization an ADaM Centre of Excellence (CoE) team was created which is a team of SME's. This team helps review existing ER and to check for any updates or improvements required.

5. Planning and Timeline

Regulatory submissions in general follow aggressive timelines, so meticulous planning and adhering to timelines is of paramount importance. It's very important to have every team member from all the functional groups to be on the same page with respect to regulatory submission timelines. It is essential to include all the tasks and deliverables when calculating timelines and estimating resources. It is again important to allow enough time for review of e-Submission package by various stakeholders including management review of various submission documents. Timelines should typically be decided and agreed early on with other functional groups and regulatory agencies as well.

6. Supplemental derivation document

This document can be utilized for the variables in the datasets that are complex and have longer derivation rules. This can be used in cases where define.xml does not supply enough ability to review, when team needs to create a table to explain few endpoints.

7. ADRG Completion Guidelines

The purpose of this document is to provide teams with a clear, concise set of instructions that facilitates the consistent development of the ADRG from the Analysis Data Reviewer's Guide Template. In addition to the ADRG Completion Guideline, ADRG examples are available as an additional reference.

8. Analysis Results Metadata (ARM)

The Analysis Results Metadata describes the key results and analysis datasets found in the analysis\adam subfolder. Teams must make sure ARM will follow the CDISC ARM Specification. Statistician is primarily responsible to identify key analyses to be included in ARM. Efforts must be made to verify all programs referenced in ARM are present in the programs folder, all the datasets referenced in ARM are present in the datasets folder. All the sponsor specific checks regarding table number, formats etc. should be in place.

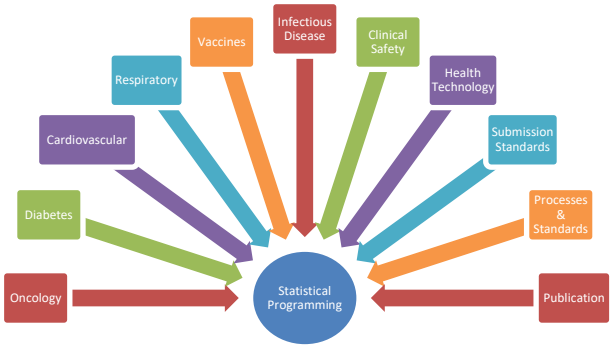
Overall Submission Package

- Cross reference across all components
  - Consistency in MedDRA version
  - CDISC SDTM/ADaM versions
  - Consistent use of definitions and acronyms across documents
  - Bookmarks & hyperlinks function and point to correct location
  - Consistent cSDRG and ADRG: data Cutoff dates, protocol versions, description, study design details, study population.

- Analysis results Metadata and SAS® programs (.txt files)
- File formats that are allowed/not allowed
- Having bookmarks and table of contents for all the PDF files

ADAM CENTER OF EXCELLENCE (COE) TEAM

There are multiple Merck statistical programming groups located in different geographical locations.



ADaM CoE team was created to achieve consistency and help study teams generate ADaM related deliverables efficiently. This team provides support on evaluating new processes, standards, and tools e.g. ADaM IG 1.1, new releases of Pinnacle 21. Team assess new SDTM version impact on ADaM implementation across statistical programming groups. Continuously support by providing review and feedback on updates to impacted ADaM related documents, execution resources and training materials. Provide support for user acceptance testing for new ADaM related utilities, new Pinnacle 21 releases and pilot projects. Provide support for reviewing ADRG new templates, completion guidelines, Study Data Technical Conformance Guide (TCG). Promote consistent implementation of ADaM within & across TAs including use of standards, processes, tools. Stay current with ADaM related submission requirements. Serve as a point of Contact for any questions, issues, feedback, suggestions for improvement.

ADaM CoE TEAM Strategy

Team collaboration can help achieve productivity and efficiency. Collaboration can be a great team building measure and can build a more creative atmosphere. This helps teams to be flexible and able to adapt swiftly to change. Here are some practical measures applied to get the most out of the ADaM CoE team, which can be used as a guidance for other organizations to create similar teams.

Assemble the Right Team
There are a set of skills that are required to have an effective team. The work is going to require a range of expertise from the team, and therefore that team must have people on it who have the various skills needed. Ideally, the team members should be self-starters.

Choose a Leader(s)
Management should play a role in identifying the leaders who have adequate experience and who can give the team accountability. If you can, get mentors to help shepherd the team and give them direction as needed.

Involve Influencers
There are some people who are great influencers. These people are well liked and respected, and they work well with others. It's easier for them to inspire other people to participate in activities. They are the perfect individuals to involve in the collaboration efforts, because they're more likely to be able to get the team engaged.

Clearly Defined Goals
It is crucial to have goals defined and in place before even assembling the team. The sooner these goals are defined, the flexible time teams have accomplishing the deliverables in highest quality.

Communication
Good communication is the hallmark of any successful project. You need to be able to articulate your needs and they must be heard and understood to move forward.

Constantly Reevaluate
To remain effective, teams must always be measuring their progress and success. Were objectives achieved? If not, why not? Adjust accordingly.

Tools to Support Team Collaboration
It is important to have a set of tools in place to make the collaboration more productive. Teams could use productivity and communication tools such as WebEx and SharePoint.

CONCLUSION

High quality submission deliverables are extremely important for regulatory review process. Establishing effective collaboration is key to success. With planning, the team can successfully achieve timelines and required compliance. Additional benefits are to achieve consistent and high-quality submission deliverables across study teams, implement new versions of ADaM standards across study teams with ease and potentially expedite the drug approval.

REFERENCES

CDISC https://www.cdisc.org/

High Quality Study Data Standards for Submission
https://www.lexjansen.com/phuse-us/2019/sa/SA03.pdf

Food and Drug Administration. "Study Data Technical Conformance Guide" Oct 2019
https://www.fda.gov/media/131872/download

PMDA (Pharmaceuticals and Medical Devices Agency) Technical Conformance Guide section 4.1.2.1
https://www.pmda.go.jp/files/000215100.pdf

CDISC "Analysis Data Model (ADaM)" version 2.1. 2009
https://www.cdisc.org/system/files/members/standard/foundational/adam/analysis_data_model_v2.1.pdf.

CDISC "Analysis Data Model Implementation Guide" Version 1.1. 2016
https://www.cdisc.org/system/files/members/standard/foundational/adam/ADaMIG_v1.1.pdf.

"Define-XML Version 2.0 Completion Guidelines" - PHUSE Wiki
https://www.phusewiki.org/docs/WorkingGroups/New%20Template%20Deliverables/Optimizing%20the%20Use%20of%20Data%20Standards/Define-XML%20Version%202.0%20Completion%20Guidelines%20.pdf

"Analysis Results Metadata for Define-XML v2" - PHUSE Wiki
https://www.phusewiki.org/docs/Denmark%202015%20SDE%20Presentations/Analysis%20Results%20Metadata%20.pdf

ACKNOWLEDGMENTS

The authors would like to thank their respective management team for their time in reviewing the paper and providing with valuable comments.

CONTACT INFORMATION

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

1. Abhilash Chimbirithy
Merck & Co, Inc.
126 E Lincoln Avenue,
Rahway, NJ 07065
732-594-6962
chimbirithy_abhilash@merck.com

2. Saigovind Chenna
Merck & Co, Inc.
351 N Sumneytown Pike,
North Wales, PA 19454
267-305-2162
saigovind.chenna@merck.com

3. Majdoub Haloui
Merck & Co, Inc.
351 N Sumneytown Pike,
North Wales, PA 19454
267-305-2629
majdoub.haloui@merck.com