

Watch Out For Blind Spots While Keeping Up With the Speed of Evolving Standards and Regulations

Srinivas Veeragoni, Bayer, Whippany, USA;
Patty Hegarty, Bayer, Whippany, USA

ABSTRACT

The new world of evolving rules, standards and regulations of clinical data, and the automation of processes for implementing these new standards are similar to the evolving features in automotive industry ranging from blind spot warnings to completely self-driving vehicles. One thing that is common and important in both these cases is to apply common sense on top of everything. To keep up with the pace of standards evolution, sponsors and CROs are trying to automate many processes in every stage from data collection to final submission. However it is important to ensure that data usability is not lost while maintaining the structural integrity of the data and compliance with standards. This paper will focus on some critical ground rules to be aware of while automating and implementing standards on clinical data.

INTRODUCTION

Clinical data standards have been evolving at a fast rate in the recent years. There are several standards and regulations expected to be conformed to, including CDISC standards, FDA/PMDA business rules & validation checks, controlled terminology, and technical conformance guides for submissions. All of these standards support consistent data collection. Operational efficiencies in the end to end data flow can be improved if you understand these standards and build systems around these. Ultimately this will help explain the traceability of the data flow and can facilitate more efficient and effective regulatory reviews. In order to keep up with the implementation of these fast evolving standards, it has become essential for sponsors and CROs to automate processes along the data flow to support data transformation and analysis, by developing metadata driven programming and validation procedures. This has also increased the scope of work for programmers, adding additional tasks such as creating metadata in relatively newer formats like XML, creating additional documentation such as cSDRG¹, ADRG², ARM³ and running and interpreting results of automated compliance checking tools like Pinnacle21®. There are some hidden dangers in all of these if we get lost in the automated world and do not pay attention to the ground rules.

Imagine you are sitting in a highly advanced self-driving car with highest ratings for safety and efficiency. Now imagine that you did not buckle your seat belt. What if another vehicle coming at a very high speed rear ends your car? The result would be very different if you had in fact buckled your seatbelt. So in this case understanding the built in safety technology, their limitations, and wearing the seat belt are ground rules to be safe in case of accidents. Similarly in clinical data automation, if data is not understood thoroughly, the standards and regulations are not interpreted properly and manual review is not done, it could lead to several issues. Among these are overblown sized datasets with hundreds of variables containing redundant information and creation of variables just for the “sake of standards” without using them appropriately or at all in the analysis. Though some level of redundancy might be built in to the standards to facilitate medical review, over redundancy could lead to loss of data value and intuitiveness and could make the data difficult to utilize for the intended analysis. Special attention should be given to the usability of the data and metadata, traceability to the point of collection, and ease of use for clinical and statistical reviewers. In this paper we will discuss some of the possible shortcomings one could encounter in standards implementation, conformance checking, and automation, and will demonstrate some precautions to overcome those shortcomings. So get set, seatbelts on, and hold on tight!

KEEPING UP WITH NEW REQUIREMENTS AND BRINGING IN NEW STANDARDS

A few years ago the pace of standards development and updates was relatively slow and scope of standards involved was smaller. In the recent years this has changed drastically with the updates being done at an exponential rate and inclusion of new standards requirements (Data exchange and Terminology) to improve the efficiency and effectiveness of receiving and evaluating regulatory submissions in different regions and safety assessments of the clinical trial data.

Figure 1 shows the timeline of SDTM and ADaM standards and compliance checks (not including Therapeutic Area User Guides (TAUGS)).

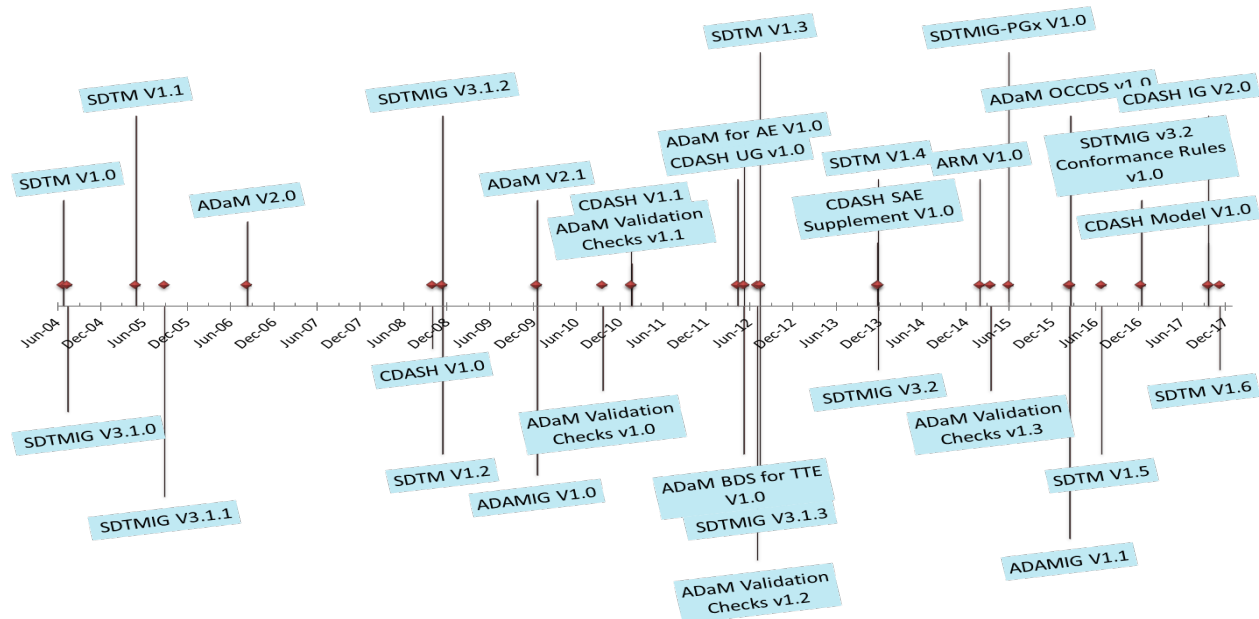


Figure 1 Timeline of SDTM and ADaM standards and compliance checks

Figure 2 shows the extensiveness of published standards (clinical) and each of these has their own version updates (not the complete list)

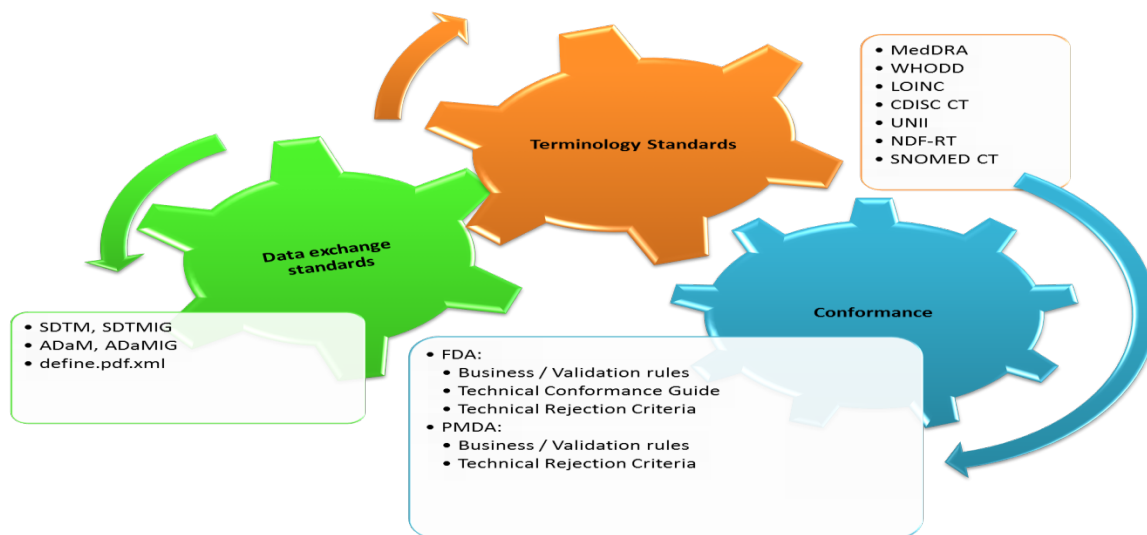


Figure 2 Extent of standards

HOW TO KEEP UP WITH THESE STANDARDS?

- Here are some of possible solutions to keep up with standards and avoid any costly consequences of non-adherence:
- **Data governance teams:** Data governance teams are dedicated teams in the sponsor companies or CROs whose primary job is to develop and maintain metadata as per the applicable industry standards. The business process owners should lead this group with appropriate representation from all the functions involved.
- **Automation:** Maintain company level metadata in a structured metadata repository which could be easily updated with the changing standards. CDISC SHARE could be useful in this process. This allows building programmatic checks to build or update the existing standards. In some cases when a version of standard is being used at the company level and if a new study needs some variables which are not available in the existing version but in a most recent version then it would be beneficial to take the variables from the new standards. This creates a hybrid version and could pose challenges while preparing Study data standardization plan, cSDRG or ADRG where the version of standards used should be specified. In these cases it would be important to get in touch with the regulatory agencies ahead of submissions and get to an agreement.
- **Trainings and Conferences:** Attending trainings and workshops like CDISC Public Training courses could help raise awareness among the programmers about the data standards and compliance. Attending webinars and conferences such as PhUSE, PharmaSUG and CDISC interchanges is another good way to get an idea of how the industry is moving forward with changing standards and may trigger ideas for efficient and optimized solutions. The other advantage of attending conferences is making connections with other industry experts and regulatory agency employees offering an opportunity to interact with them in person which might not be possible in a typical study setting.
- **Volunteering in Industry Groups:** Organizations like CDISC, PhUSE and TransCelerate BioPharma are some examples of industry groups which rely on volunteers and/or sponsorships from the industry who could participate in regular discussions and contribute to the development of standards to streamline clinical research. There are several working groups within these organizations such as PhUSE working groups that focus on standards implementation and help advance the standards and close the gaps. Individual contributors could join and contribute based on their interests and skill sets. This works as a two way learning i.e. the working groups benefit from individual expertise and the contributors learn from the rest of the cross industry teams. These learnings if brought back to the companies are useful in designing and maintaining the internal systems and processes on par with the evolving standards.

DIFFERENCE BETWEEN HEALTH AUTHORITY AGENCY BUSINESS/ VALIDATION RULES AND CDISC COMPLIANCE CHECKS

Health Authorities publish information that is intended to help sponsor companies prepare high quality, complete packages of electronic data and supportive documentation. This information covers submission of electronic data, acceptable data formats, and validation checks that are run on submitted data. Currently the US FDA, Japanese PMDA, and Chinese cFDA require submission of electronic data at the individual patient level for clinical studies. The EMA does not yet require individual patient data to be submitted electronically. There are many common expectations between these agencies as to the format and content of the data submitted, as well as some key differences. There are commercially available tools and automated checks that help to verify conformance of the structure and content of electronic data with industry standards capable of identifying issues in the submission package. At the same time, it is also important to have a good understanding of what might not be covered in these automated checks. Additional topics such as data quality, traceability, and handling of legacy data are other factors that might impact the overall quality of the submission and experience for the reviewer. You should educate yourself on all the standards and guidance and have a robust preparation and review process in place to not get lost in the automation.

Table 1 presents an overview of similarities and differences of the requirements from FDA and PMDA:

Topic	FDA	PMDA
General Requirements	FDA requires the submission of individual patient data in electronic format from clinical and non-clinical studies to support review and approval of drug applications.	PMDA requires the submission of individual patient data in electronic format from clinical studies to support review and approval of drug applications.
General Documents	Technical Rejection Criteria ⁴ is a published list of rules outlining required datasets and documentation files, along with technical instructions for submission for all clinical and nonclinical studies starting after December 17, 2016*. The FDA may refuse to file a submission where data does not conform to standards specified in the FDA Data Standards Catalog. (*please NOTE: submission of TS.xpt required for all studies even those conducted with start dates prior to December 17, 2016.)	<p>The Basic Principles on Electronic Submission of Study Data for New Drug Applications¹⁰ contains general principles supporting the PMDA's efforts to collect and analyze clinical study data in electronic format as they relate to the review of new drug applications. The document outlines the specific standardized data formats that are expected, as well as expectations on other supportive files including documentation and analysis program files.</p> <p>Notification on Practical Operations of Electronic Study Data Submissions Guide¹¹ is a document providing detailed info on the preparation and submission of electronic patient data and documents to support drug applications. The guide provides detailed information on the following topics:</p> <ul style="list-style-type: none"> • technical details on how to submit electronic data to PMDA through the electronic gateway, including acceptable timing of submissions • expected conformance to CDISC/SDTM and ADaM data standards • supportive documentation & program files to be submitted • validation rules to be followed i.e., use of the OpenCDISC Enterprise Tool [renamed to Pinnacle21® Enterprise Tool], and statements on consequences if non-compliant data is submitted • expectations on consultations between the PMDA and sponsor company controlled terminology and coding dictionaries
Study Data Technical Conformance Guide	The Study Data Technical Conformance Guide ⁵ is a guidance document providing detailed info on the preparation and submission of electronic patient data and documents to support drug applications.	The Technical Conformance Guide on Electronic Study Data Submissions ¹² is a document outlining the process for how and when to submit electronic data to the PMDA, including the timing and

	<p>The guide provides detailed information on the following topics:</p> <ul style="list-style-type: none"> • supportive documentation & program files to be submitted • accepted exchange formats for clinical and non-clinical data • expectations around use of industry standards (e.g., CDISC and Therapeutic Area Standards) • controlled terminology including coding dictionaries • technical details for placement and tagging in the eCTD backbone • data validation checks and traceability between data models • handling of legacy data • links to other resources with helpful information. 	<p>type of consultations that should take place between the sponsor company and PMDA. The document also outlines the technical details for placement and tagging of files in the eCTD backbone, as well the Conformance Rules and Validation Rules that are applied to checking the data in SDTM and ADaM formats. The document details the expected submission of Analysis Results Metadata (ARM) to connect the data to the results and other supportive documentation such as the annotated CRF, reviewer's guides, software programs, etc. There is also a lengthy section on the submission of pharmacokinetic data including population modeling and analysis, and simulations, and a section on submission of Japanese text for certain free text result fields.</p>
<p>Business Rules, Conformance Rules, Validation Rules</p>	<p>Business Rules⁶, Validation Rules⁷, and Conformance Rules⁸ are a series of specific structural and content checks on standardized data defined by FDA, CDISC, and/or outside vendors e.g., Pinnacle 21, meant to ensure data are compliant and will support review. These rules check for the presence/absence of datasets and/or variables, use of controlled terminology, and perform content and consistency checks on specific variables. The Business Rules state the check in English language, while the Conformance and Validator rules give more detailed information on the specific domain names and variable names to be checked, based on version of the data standard (e.g., SDTM, SEND, ADaM).</p> <p>Understanding both the model (SDTM or ADaM) and the respective IG is important for correct implementation. Prioritization of conformance checks should be based on hierarchy relevant to submission¹⁸.</p>	<p>Conformance Rules, Validation Rules¹³: In addition to the previously described CDISC Conformance Rules, PMDA has published a set of Study Data Validation Rules for SDTM, ADaM, and the define.xml. Each of these rules is assigned a severity level: warning, error, or reject.</p>
<p>Data Standards Catalog</p>	<p>The FDA Data Standards Catalog⁹ outlines the accepted standard data formats for SDTM, ADaM, SEND, the define.xml, and controlled terminologies among other data standards and file types, along with dates that support for various versions begins and ends.</p>	<p>The PMDA Data Standards Catalog¹⁴ outlines the accepted standard data formats for SDTM, ADaM, define.xml, and controlled terminologies, along with dates that support for various versions begins and ends.</p>

Table 1 overview of similarities and differences of the requirements from FDA and PMDA

NOTE: It is advisable to discuss plans for submission of electronic data in detail with the Health Authority prior to submitting a drug application, including an assessment of data quality and conformance to standards.

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.

WHERE ARE SOME BLIND SPOTS?

COMPLIANCE VERSUS CONTENT CHECKS

As discussed above one of the solutions to keep up with the fast changing standards is automation. Automation allows checking most of the compliance with standards programmatically by using tools like Pinnacle21® or in-house developed quality check tools. However, as mentioned in the CDISC compliance check documents, not all checks can be programmed. For example, many ADaM variables are conditionally required (required if a condition is true), but some conditions are not testable by a software algorithm. In addition you need to make sure all the guidelines in the regulatory guidance are followed. It is important to understand the difference between compliance (structural checks) vs content checks (usability). You should try to establish a process in place to ensure the usability of the data for the intended analysis. Let's think this in terms of a double programming concept using SAS® procedures like COMPARE which gives the exact comparison between two outputs. We know that the PROC COMPARE output of "NOTE: No unequal values were found. All values compared are exactly equal." does not always mean that the output is correct or as expected. The outputs might be matching even if the output is not usable for the intended purpose due to various reasons such as specifications being wrong, interpretation of specifications by both the programmers could be wrong, the input data to the program might have been wrong or both the outputs could be blank. So it is always important to have a manual check for specific data that cannot be checked programmatically and make sure that the results are as expected. It is equally important to understand and interpret the results in addition to producing them.

OVERBLOWN SIZE OF DATASETS

Incorrect interpretation of standards and efforts to make the data analysis ready could sometimes lead to overblown size of datasets with unnecessary variables. This makes the datasets huge and could become burdensome for reviewers to distinguish important data from the unimportant data. Some of the possible reasons for adding these unnecessary variables could be populating permissible or conditional variables even if they are blank or deriving them if not needed for analysis. As per SDTMIG¹⁵ the sponsor can decide whether a Permissible variable should be included as a column when all values for that variable are null. The sponsor does not have the discretion to exclude permissible variables when they contain data. For example FASFL and ITTFL are used interchangeably by some sponsors. In this case choosing one of these population flag variables and using it consistently across all analyses is recommended instead of populating both if their definitions are same. As per ADaM model¹⁷ "The inclusion of too many extraneous variables (i.e., variables not needed to support analyses) makes it more difficult for users to find important variables and can impede clear and concise communication."

Table 3.2.3 ADSL Population Indicator Variables

Variable Name	Variable Label	Type	Codelist/ Controlled Terms	Core	CDISC Notes
FASFL	Full Analysis Set Population Flag	Char	Y, N	Cond	These flags identify whether or not the subject is included in the specified population. A minimum of one subject-level population flag variable is required in ADSL.
SAFFL	Safety Population Flag	Char	Y, N	Cond	Not all of the indicators listed here need to be included in ADSL. As stated in Section 3.1.4, Item 2, only those indicators corresponding to populations defined in the statistical analysis plan or populations used as a basis for analysis need be included in ADSL.
ITTFL	Intent-To-Treat Population Flag	Char	Y, N	Cond	This list of flags is not meant to be all-inclusive. Additional population flags may be added.
PPROTFL	Per-Protocol Population Flag	Char	Y, N	Cond	The values of subject-level population flags cannot be blank. If a flag is used, the corresponding numeric version (*FN, where 0=no and 1=yes) of the population flag can also be included. Please also refer to Section 3.1.4.
COMPLFL	Completers Population Flag	Char	Y, N	Cond	

Display 1: ADSL Population Indicator Variables AS DEFINED IN ADAMIG V 1.1¹⁵

One other common reason could be deriving variables with redundant information due to wrong interpretation of standards. One example could be the use of relative timing variables such as --STRF and --ENRF. In general these variables are collected on the CRF and presented in SDTM domains. However in some cases sponsors may wish to derive these variables for analysis or reporting purposes in which case imputation rules should be considered. If these additional variables are derived in analysis datasets without any need for analysis or reporting then this would be redundant information. This could happen mostly when the company has a global metadata with all possible variables and while adapting to the individual studies if the proper attention is not paid to distinguish mandatory vs optional variables. Here is one of the fundamental principles of ADaM datasets: "It is not necessary to collate data into "analysis-ready" datasets solely to support data listings or other non-analytical displays" (Section 2.1 of ADaMIG V1.1¹⁵). Another good example for this case as shown in Table 2 could be a misunderstanding of the requirement of code and decode pairs for every possible variable which might not add any value and be redundant information. These issues could be avoided by keeping yourself up to date with standards and clarifying anything in doubt with the company data governance teams, external industry blogs and /or by checking with the regulatory agencies during the pre-submission meetings.

Domain	Variable	Value	Derivation	Comment
ADSL	REGION	North America	stratification factor collected on CRF QVAL where SUPPDM.QNAM="REGION"	only character value collected on CRF
ADSL	REGIONN	1	SUPPDM.QNAM="REGION" converted to numeric	character value converted to numeric for analysis
ADSL	REGIONC	North America	decode of ADSL.REGIONN	Decode character value created from numeric value REGIONN. This is redundant information with REGION variable

Table 2: Example of redundant variables due to lack of clarity about code and decode variables

IMPORTANT VARIABLES WITH INCORRECT INFORMATION

Another shortcoming while trying to adhere to the standards without complete knowledge could be populating important variables with incorrect information. For example deriving EPOCH can be challenging based on incomplete or missing source data. Care must be taken to ensure that the derived period in the epoch is accurate and could be used for any analysis. Users should keep in mind that some of the key variables such as EPOCH which are expected per documents like technical conformance guide could be important for the regulatory reviewers to reproduce the analysis using standard scripts, even if these variables are not used in the analysis and reporting. It is very important that the results presented in the Clinical Study Report be traceable back to the original data elements as they were collected and represented in SDTM.

SUBJIDN	SESEQ	ETCD	EPOCH	SESTDTC	SEENDTC	Comment
101	1	SCRN	Screening	2017-10-31	2017-11-22	
	2	TRT	Treatment	2017-11-22	2018-01-27	
	3	SAFFU	Safety FU	2018-01-28	2018-01-27	Data issue SESTDTC > SEENDTC due to Last Treatment date (2018-01-27) = Last Visit Date. This data issue may be caused by the subject not having a safety follow-up epoch or the date of safety follow-up visit might be wrong.

Table 3: Example of an important variable EPOCH with incorrect information (as a result of wrong SEENDTC).

DOCUMENTATION

Data specifications must be provided in define.xml format as well as a cSDRG (for SDTM) and an ADRG (for ADaM). While building the metadata repositories and automating programming based on the metadata, it could be enticing to lean towards making the specifications more algorithmic for the macros to read in and execute the programs. However, it is important to know that these documents are not only for programming and validation but also for statistical reviewers and sometimes even medical reviewers who wish to reproduce or understand the analysis in depth. To facilitate the review process, analysts could tend to choose to document all descriptive parts in the reviewer’s guides which could lead to additional problems like detaching the actual derivation logic from the source of the variable in define file and putting it separately in reviewer’s guide. This could also add additional burden on the reviewer and could sometimes delay the review process. An insufficiently documented define file is a common deficiency that reviewers have noted. So you should consider making the ADS specifications (define.xml) to be more intuitive to understand how the variables are derived, the source data being used along with any applicable controlled terminology, while trying to maintain its usability for automation. The reviewer’s guides should be used to focus on special consideration or directions to help facilitate review, define any anomalies or data discrepancies and assist the reviewer understand the relationship between the datasets and the CSR.

Variables used in the analysis	where AE.EPOCH in (“Treatment Epoch”, “Safety Follow-up Epoch”)	ADAE.TRTEMFL = “Y”
How the variables differ in their content	Treatment epoch = First Dose Taken to last intake of study medication Safety Follow-up Epoch = Day after last intake of study medication to 30 + 4 days after last intake of study medication (according to protocol)	Defined as any event arising or worsening after start of study drug administration until 30 days after the last study drug intake
Number of subjects with Treatment emergent AEs	201	200

Table 4: Example of documenting some known differences using different analysis variables in ADRG

IMPROPER SETUP AND USE OF VALIDATION TOOLS

As discussed in the earlier sections there are commercial tools like Pinnacle21 or sponsor developed tools to check compliance. The users of these tools need to understand how these tools are programmed and need to be able to make sense of the outputs of the tools. While relying on automation tools, it is very important that the tools are configured and setup properly for them to work properly. You should ensure that the data that is being passed is the right source data; the standards being referenced for the checks are consistent with the source data and those referred in SDSP, cSDRG, ADRG and briefing documents. If the environment is not setup appropriately you might not get the same validation findings as those of the regulatory agencies get when they run similar tools.

CONSIDERATIONS FOR SAME DAY FILING

Same day filing refers to sponsors filing NDA to multiple regulatory agencies like FDA, EMA, PMDA, cFDA etc. at the same time. In addition to understanding different regulatory data standard requirements (as explained in Table 1 in this paper) and preparing the data and reporting packages accordingly there might be few additional things which might not be clear from the standard documents. One example of this case is shown below. These kinds of blind spots could be avoided if proper consultations with the regulatory agencies are made at appropriate time during submission preparation some of which are explained in the next section.

Regulatory agency	Dataset	version of WHODD
FDA	CM ADCM	version used during study conduct Ex: "WHODD September 2005 with updates"
PMDA	CM (one dataset for each version of WHODD)	version used during study conduct Ex: "WHODD September 2005 with updates"
	ADCM (one dataset for each version of WHODD)	the latest version at the time of submission "WHO-DD March 2016"

Table 5: Difference in coding dictionaries to be used as requested by FDA and PMDA

DATA EXCHANGE SYSTEMS:

During submission process you have to deal with several data exchange systems both within the sponsor environments and the regulatory agency gateways. It is important to cross check the information uploaded to sponsor document management system (or agency gateways) versus content in the internal statistical programming environment. There are chances that the upload could have been partial or the data might have been accidentally transformed or corrupted during the upload process. So it might be useful to download the uploaded files back out of the document management system and do a proc compare or check the size of the uploaded file with the source file.

IMPORTANT POINTS TO CONSIDER FOR A WELL PREPARED AND COMPLETE SUBMISSION PACKAGE

1. IMPORTANCE OF ANALYSIS RESULTS METADATA (ARM)

Analysis Results Metadata provides traceability from results in a statistical display to the data in the analysis datasets. It facilitates documentation and reproduction of the analysis results. CDISC standards give us a standard to organize our data and ARM tells us how this data is utilized in a standard format which is important part of the traceability emphasized by regulatory agencies like FDA. This is not needed for every analysis in a submission but is important for reproducing key tables by regulatory reviewers.

2. IMPORTANCE OF SUBMISSION OF PROGRAMS

As per the Study Data Technical Conformance Guide (SDTCG) sponsors should provide the software programs used to create all ADaM datasets and generate tables and figures associated with primary and secondary efficacy analyses. The main purpose of requesting the submission of these programs is to understand how the respective analyses were created and to confirm the analysis algorithms. Sponsors should be aware that this helps the reviewers understand the derivation of critical variables, makes the process transparent, and could speed up the review process.

3. IMPORTANCE OF MEETINGS AND COMMUNICATIONS WITH REGULATORY AGENCIES

Regulatory meetings plan an important role for successful and efficient submissions. Some of the common meetings with agencies are described below.

I. MEETINGS WITH FDA:

The list below in Figure 3 below is to give an overview of the meetings with FDA for agreements on electronic data and might not be a comprehensive list.

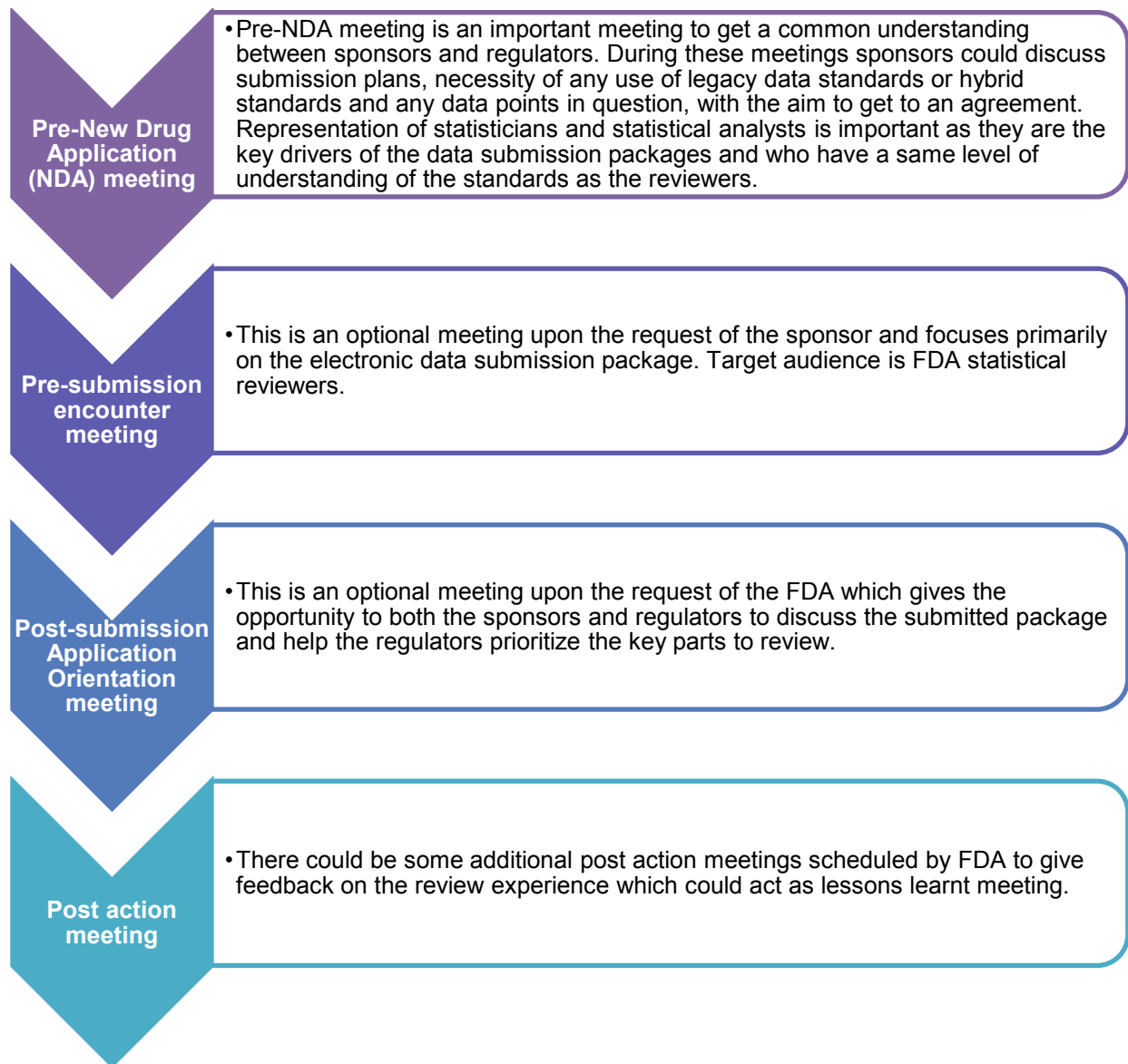


Figure 3: Some meetings with FDA

II. SOME IMPORTANT DOCUMENTS FOR COMMUNICATION WITH FDA:

The list below is to give an overview of some important documents that are used for communicating with FDA and might not be a comprehensive list of all documents.

a. Study Data Standardization Plan (SDSP)

SDSP contains information about the standards used in the compound such as versions of CDISC standards used, controlled terminology, dictionaries etc. for both non-clinical and clinical studies. This is an important document for discussions with FDA and documenting high level decisions and agreements. This document is a living document which could be updated over the life cycle of the compound.

b. Briefing package

Briefing package is another document where sponsor could summarize the submission plan and get clarifications and agreements on the structure and content of the electronic data packages for the submission.

III. MEETINGS WITH PMDA:

PMDA requires at least two meetings before submission of a drug application as shown in Figure 4.

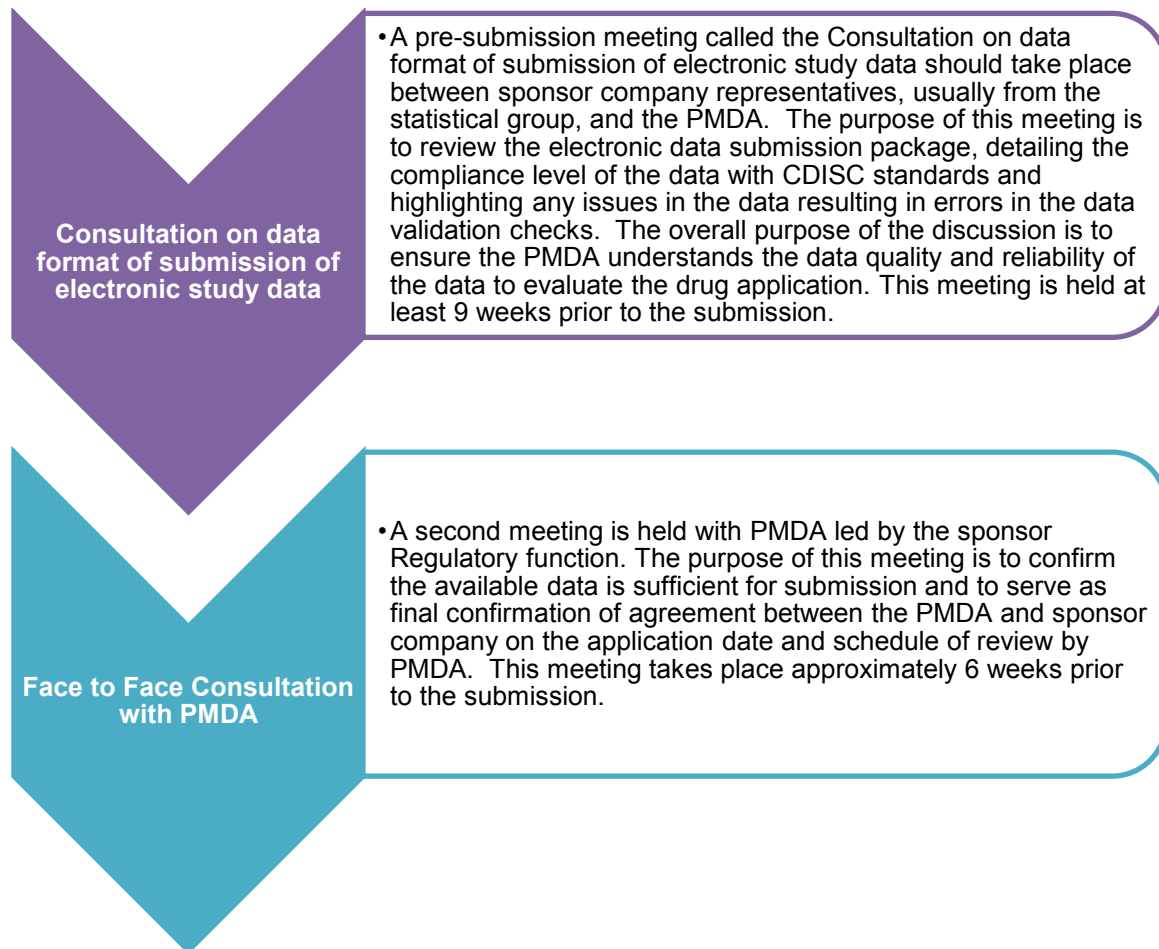


Figure 4: Some meetings with PMDA

IV. SOME IMPORTANT DOCUMENTS FOR COMMUNICATION WITH PMDA:

The list below is to give an overview of some important documents that are used for communicating with PMDA and might not be a comprehensive list of all documents.

a. Material required for Face to Face Consultation with PMDA

As part of the Face to Face meeting with PMDA, the sponsor should submit information on the background of the drug product under development, including other existing treatments and expected benefits of this new drug. Draft protocols, the latest Investigator's Brochure, development status in other countries, and other references should also be submitted as part of the information package for the Face to Face Consultation Meeting.

4. EFFICIENTLY HANDLING INFORMATION REQUESTS FROM AGENCIES

It is common during the review process that the reviewers might come back to the sponsors and ask for additional information or clarification of the submitted package which might be needed to complete the review process. Sponsors could avoid some of the Information Requests (IR) if you have proactively submitted the entire package in standard format with complete and usable content such as submitting programs for key efficacy without any macro code, ARM etc. However there could be some additional IRs which asks for data in non-standard format to support specific review questions.

ALIGNING BETWEEN STATISTICAL FUNCTION AND OTHER FUNCTIONS INVOLVED IN SUBMISSION

Representatives from the Statistical function are the subject matter experts for electronic data topics as it relates to clinical study data used for analysis. They must closely collaborate with other contributing functions such as Pharmacokinetics (PK), Integrated Analysis (IA), Medical Writing, Clinical, Publishing, and Regulatory to ensure the preparation of a comprehensive, high quality, compliant submission package. Early and upfront planning is required to align with these functions, well in advance of any Regulatory Agency meeting(s) to discuss the submission contents e.g., pre-NDA meeting with FDA or Consultations with PMDA. Potential blind spots in cross-functional interactions may include the following:

- Pharmacokinetics (PK) function: Consistency between the clinical data prepared by the Statistical function and data presented by the PK function is of utmost importance. Demographic, drug exposure, and individual PK concentration data, among others may be presented in data and reports by both groups.
- Integrated Analysis (IA) function: Early planning between IA programmers and the single study programmers will ensure consistency between individual study results and integrated summary results. Proper compliance with standards in individual studies will help to ensure efficiencies in data pooling and integrated analysis. All the standards applicable to study data are also applicable to IA data.
- Medical Writing & Clinical functions: The Medical Writing and Clinical functions are key customers of the data and reports produced by the Statistical function. Close collaboration is required to ensure necessary displays are available for presentation in clinical study and integrated summary reports.
- Publishing function: Here is it important that the Statistical function act as subject matter experts on the requirements around submission of electronic data packages, as the Publishing function likely does not necessarily have a deep understanding of such requirements. Although Publishing ultimately compiles the submission package, they may not have necessary familiarity with clinical data structures and might not have tools (e.g., SAS® Viewer) to open and review the data as part of the validation of the eCTD submission package.
- Sponsor Regulatory function: Similar to Publishing, Regulatory might only have a high level of understanding of the clinical data and relevant standards. It is important for the Statistical function to help them understand and interpret the technical conformance guidance from the regulatory agencies.

STATISTICAL ANALYSIS: INDUSTRY AND REGULATORY INTERACTIONS

As the requirements for submission of electronic data are complex and continuing to evolve, it is suggested to have a dedicated point of contact within the Statistical function familiar with regulatory submission requirements and industry developments within the topic of electronic data. This point of contact should mentor and provide advice on the preparation of electronic submission packages to those who are performing operational work on the compound in the Statistical function. The point of contact can also support the education of other functions mentioned in the previous section on overall submission data standards and compliance processes. By providing dedicated support for partnering functions, the Statistical group will be seen as a trusted function for all data relevant topics and included in key interactions with Regulatory Agencies which would help to address some of the blind spots discussed in this paper.

CONCLUSION

Clinical trials and drug approval processes are highly complex, and it could be challenging for reviewers to navigate through the overwhelming amount of data submitted to them in order to efficiently review drug applications. You have seen the various industry standards that are available and continuing to evolve at an exponential rate, as well as the expectations from health authorities to comply with these standards. It

may be tempting to automate many processes in every stage from data collection to final submission to keep up with these fast evolving standards; however, there are hidden dangers in automated processes if you blindly apply algorithms and tools without paying attention to the usability of the resulting outputs to support review and analysis. You have seen the importance of being compliant with standards while maintaining usability of the content, and have also seen the consequences of not completely understanding the standards or how best to apply them. To address these blind spots, this paper has suggested several possible solutions ranging from creation of comprehensive documentation, participating in meetings with health authorities to discuss data relevant topics, and taking precautions while automating. You have also seen how to influence the development and maintenance of standards by participating in industry wide forums and the importance of having dedicated personnel to implement standards at your company. Remember: do not solely rely on a self-driving car to deliver you to your destination. Take into account other helpful technological advances such as blind spot warning systems and GPS. Remember also to keep your eyes on the road and be on the lookout through the windshield, side mirrors, and rear-view mirrors! Use all available tools to reach your destination safely and on-time!

REFERENCES

¹ PhUSE “Study Data Reviewer’s Guide” available at:

http://www.phusewiki.org/wiki/index.php?title=Study_Data_Reviewer%27s_Guide

² PhUSE “Analysis Data Reviewer’s Guide” available at:

http://www.phusewiki.org/wiki/index.php?title=Analysis_Data_Reviewer%27s_Guide

³ CDISC “Analysis Results Metadata (ARM) v1.0 for Define-XML v2.0”. 2015. Available at:

<https://www.cdisc.org/standards/foundational/adam>

⁴ FDA “Technical Rejection Criteria” available at:

<https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM523539.pdf>

⁵ FDA “Study Data Technical Conformance Guide” Oct2017 available at:

<https://www.fda.gov/downloads/forindustry/datastandards/studydatastandards/ucm384744.pdf>

⁶ FDA “FDA Business Rules v1.3” Dec2017 available at:

www.fda.gov/downloads/forindustry/datastandards/studydatastandards/ucm546467.xlsx

⁷ FDA “FDA Validator Rules v1.2” Dec2017 available at:

www.fda.gov/downloads/forindustry/datastandards/studydatastandards/ucm546468.xlsx

⁸ CDISC “SDTMIG v3.2 Conformance Rules v1.0” Jan2017 available at:

<https://www.cdisc.org/system/files/members/standard/foundational/sdtmig/SDTMIG%20v3.2%20Conformance%20Rules%20v1.0.xlsx>

⁹ FDA “Study Data Standards Catalog” Oct2017 available at:

<https://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM340684.xlsx>

PMDA “Overview of Reviews of Drug Applications” available at:

<https://www.pmda.go.jp/english/review-services/reviews/0001.html>

¹⁰ PMDA “Basic Principles on Electronic Submission of Study Data for New Drug Applications” Jun2014 available at: <https://www.pmda.go.jp/files/000153708.pdf>

¹¹ PMDA “Notification on Practical Operations of Electronic Study Data Submissions” Jul2015 available at: <https://www.pmda.go.jp/files/000206451.pdf>

PMDA “Statements on Review with Electronic Data” <date> available at:

<https://www.pmda.go.jp/english/review-services/reviews/advanced-efforts/0002.html>

¹² PMDA “Technical Conformance Guide on Electronic Study Data Submissions” Jul2015 available at: <https://www.pmda.go.jp/files/000206449.pdf>

¹³ PMDA “Study Data Validation Rules” available at: <https://www.pmda.go.jp/english/review-services/reviews/advanced-efforts/0002.html>

¹⁴ PMDA “Data Standards Catalog” available at: <https://www.pmda.go.jp/english/review-services/reviews/advanced-efforts/0002.html>

¹⁵ CDISC “Study Data Tabulation Model Implementation Guide V3.2” 26NOV2013 available at: https://www.cdisc.org/system/files/members/standard/foundational/sdtmig/sdtmig_20v3.2_20noportfolio.pdf

¹⁶ CDISC “Analysis Data Model Implementation Guide V1.1” 02DEC2016 available at: https://www.cdisc.org/system/files/members/standard/adam/ADaMIG_v1.1.pdf

¹⁷ CDISC “CDISC Analysis Data Model Version 2.1” 17DEC2009 available at https://www.cdisc.org/system/files/members/standard/analysis_data_model_v2.1.pdf

PMDA “Overview of Consultation System in Japan” 22MAR2012 available at: <https://www.pmda.go.jp/files/000157641.pdf>

¹⁸ Michael Wise, Soumya Rajesh, and Kapila Patel. 2018. “SDTM: Model or Implementation Guide?” PhUSE Computational Science Symposium, Silver Spring, MD.

Available at: <http://www.phusewiki.org/docs/2018%20US%20CSS/posters/PP12.pdf>

ACKNOWLEDGMENTS

We would like to thank Sascha Ahrweiler, Ankur Mathur and Tripti Bhaskaran for their review and valuable comments.

CONTACT INFORMATION

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

Srinivas Veeragoni
Bayer US LLC
Srinivas.Veeragoni@Bayer.com

Patty Hegarty
Bayer US LLC
Patricia.Hegarty@bayer.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.