Paper DH04

Data Compliance: Hidden Benefits

Priyanka Adusumilli, Accenture, Berwyn, USA
Priyank Patel, Accenture, Berwyn, USA

ABSTRACT

The pharmaceutical industry is currently spending a large amount of money updating the CDISC SDTM submission packages according to FDA standards. Once study data is deemed ready for submission by a vendor, most sponsors expect the compliance checks are completed, merely to update the data for that study. However, those same findings from the compliance checks can be used to identify bigger issues with the overall collection strategy. John Powell once said, “The only real mistake is the one from which we learn nothing.” Addressing the issues within the individual study for compliance checks solves the immediate problem, until the next study comes along and the same findings are prevalent. Being able to learn from these findings and use them in a positive manner beyond the individual study can help create efficiency and provide a more transparent data transformation strategy that will be in line with the FDA’s expectations.

INTRODUCTION

The FDA emphasizes the submission of standardized data for drug applications for numerous reasons, mainly for ease of review and traceability. The source data is standardized into the Study Data Tabulation Model(SDTM) for clinical studies. Non-clinical data must be submitted using the Standards for Exchange of Nonclinical Data (SEND). The implementation guide for the SDTM provides guidance on what domains and variables can be used within the SDTM model. This document also provides examples on how the data should be represented. The SDTM data is then used to produce Analysis Data Model (ADaM) and Tables, Figures and Listings(TFLs). Like the SDTM Implementation guide, there is a similar guidance document for ADaM.

In addition to the implementation guides, the Study Data Technical Conformance Guide is also a guiding document in data submission to the FDA. This document provides overall guidance on the submission strategy and key items to consider. It contains rules regarding the electronic submission of the data, including the data, define and other publishing documents. Some items covered in the Technical Conformance Guide include dataset size, variable length, reviewer’s guide expectations, etc.

Currently, the FDA does not endorse a compliance check engine, however, the industry standard is using Pinnacle 21. Pinnacle 21 incorporates all the rules from the different implementation guides for different data formats as well as guidance from the Study Data Technical Conformance Guide. There are checks that come up as errors or warning based on the severity of the issue. The issues that can be fixed should be fixed and those that cannot should be included in the reviewer’s guide with a rationale provided.

CURRENT PROCESS

COMPLIANCE CHECKS

The compliance report produces all types of errors, warnings and notices. Being able to identify the difference between data errors and transformation errors is crucial in determining out how to resolve the issue. Data errors are source data issues where the data was collected as such and an update to the standardized data isn’t possible without still maintaining the integrity of the standardization process. For example, a protocol deviation that was recorded prior to the subject starting in a study. In cases like this, a rationale would be provided in the reviewer’s guide.

Transformation errors are errors due to the process of standardizing data. For example, required variables that are not populated. The required variables are expected based on the Implementation guide and serve a specific purpose for the FDA reviewers. This error should be fixed in the standardized data. It is worth noting that every so often, there are some false issues that come up. For example, variable labels that are suggested in the implementation guide might be too long for the limits of that variable. A simple rationale would suffice in explaining why this issue is showing up.

COMMON ISSUES IN DATA TRANSFORMATION

The data collection is a big part of the process as it determines the quality of the data to be submitted. If the data collection is thorough in its’ emphasis on the outcomes outlined by the protocol, there is minimal chance that there will be much trouble with the standardization process. There are many cases where the necessary data was not

collected completely, or the integrity was questioned, rending it useless. Having a clear outline of the expectations, as laid out in the protocol, are crucial to developing a study and data management plan.

The current process is developing the standardized data and then running the pinnacle report to check for any of the issues, updating the data, re-running pinnacle checks, and repeat until all issues are addressed, whether fixed or rationale provided. This creates quite a churn and waste of time and resources. Most of the compliance errors can be prevented with the right structure/system in place to develop domains.

In addition to compliance checks, there are common issues that the FDA found that can be addressed prior to submission. During a FDA webinar last year, dealing with SEND submissions in particular, the most common issues based off all SEND submissions in the year 2017 were laid out. The same could be translated to other standardized data in a general sense. One of the issues discussed is the lack of an SE(Subject Elements) domain, which would therein not have the EPOCH variable populated. Without the EPOCH variable, it's difficult for the reviewer to determine during which part of the trial a finding or result is from. There would be a corresponding compliance error in pinnacle.

Another common issue is lack of proper standardization of the lab data. This can be applied to multiple domains in terms of standardizing the units and tests. Without the standardization, different data vendors can call the same test different names and collect the results in different units. Using controlled terminology, the standardization allows easy compare and review of the data submitted. Even though this compliance issue would show up in pinnacle as a warning, it would be prudent to pay attention to it as it was called out specifically. It is highly recommended to conform to the controlled terminology.

WHAT CAN BE DONE?

The first part of any solution that is explored must be knowledge. Knowledge of the requirements, knowledge of the guidance documents, and knowledge of compliance checks. This can be directly translated to an easy solution of training. Training would need to incorporate all the items mentioned in order to make this as thorough as possible. The requirements and guidance documents would include the standardization models, implementation guides and the study data technical conformance guide. There is an abundance of information related to the submission of standardized data. Learning how to read the compliance report and understanding the compliance checks would also go a long way to knowing which issues can be updated and which need to have a detailed rationale in the reviewer's guide. In addition, learning how to use the controlled terminology is crucial in reducing many of the compliance checks. This allows true standardized data for submission.

Armed with the knowledge, the other solutions can be tackled. One approach that would set the whole study up for success is to format the data collection to fit the standardized data. For example, it would help to have a standardized data entry for date of birth or other dates collected. There could be multiple ways of collecting the same date information, but it needs to be standardized to the same format in order to submit the data to the FDA. The added confusion of whether the month or day are being switched depending on the country the date is collected in. If the dates were standardized, this would eliminate the need to go through the hassle of figuring out the code for conforming the data to CDISC standard down the line when the SDTM is developed. Something as minute as this would save time in the long run. The same could be done for lab tests, units and carried forward to other domains or variables. When a sponsor has multiple studies that are being set up, this time could add up and would provide the consistency that is deeply lacking in pharmaceutical data.

The next solution would be to have a platform that could work with raw data and provide a structure for the standardized data. Due to the various types of raw data structure, at this time, this would still be a manual effort in terms of determining where the raw data would belong. This platform could have table shells that show which variables are required, which variables are permissible, which would help the data developer to better understand the requirements. Right off the bat, a few issues can be resolved; ensuring all required variables are populated, extraneous variables added into domain that are not part of the standardized data model, removing permissible variables that are null. Another component of this platform would be to automatically incorporate the variable lengths, variable labels, domain labels, etc. At a minimum, the platform could check if some of these issues persist and warn the user of the issues in order to update immediately rather than waiting until the compliance checks are run and then updating the data.

WHY CHANGE?

Providing a clean submission to the FDA will allow for faster and easier review with fewer questions. If there are many issues impacting the data quality, it questions the integrity of the data. In addition, if there are key items that are missing from the submission, the FDA reviewer could ask the sponsor to provide this information, which would be quite costly and needing to essentially restart the process in order to incorporate the new data. Had the issue been caught earlier with compliance checks, it could have been addressed at the same time as the data transformation. The number and type of issues that are quite prevalent within a particular sponsor's trials could suggest other routes

for improvement in data collection. Something as simple as a required variable not being collected could allow the sponsor to bring that experience to future trials and avoid this issue moving forward.
FDA reviewers have previously noted that even trials within the same compound might not be consistent depending on the data collection and the data standardization. In using any of the solutions, there could be a profound improvement in consistency of the data to allow for easier review and comparison. The compliance checks are useful in understanding what the common issues are between all studies. Fixing it in one study can provide a path forward for all others.

In addition to consistency, the traceability of a package could be improved overall. This is a major issue in terms of being able to trace anything in the study report back to the analysis data, to the SDTM/SEND data, to the source data. That level of traceability instills confidence in the results and the data transformation. If there is something in the data that did not make it into the report, that would raise red flags and could question the integrity of the submission, depending on the severity.

Imagine a sponsor being able to use the knowledge of the entire process and requirements and be able to incorporate the data collection and the platform for data transformation. The seamless way data can be standardized as it would be collected to go into the required domains or variables. The platform would then produce the data based on the requirements laid out by the FDA. The automation of the platform would eliminate a lot of human error. If certain items were pre-populated such as domain labels, this would reduce the effort needed to reproduce this with every study. The efficiency in the entire process would allow for a quicker turnaround in the standardized data, allowing a decrease in production time needed for the data transformation.

CONCLUSION

All the approaches above would allow the sponsor to get to their submission faster, resulting in monetary savings. The overall quality of the standardized data would improve drastically because of the traceability and consistency of the data. This also leads to a better submission strategy and allows for more meaningful discussions with the FDA reviewer. Understanding the requirements and the shortcomings is the first part of the process. Status quo cannot sustain within this industry.

CONTACT INFORMATION

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

Priyanka Adusumilli
Accenture
1160 W. Swedesford Rd.
Berwyn / 19312
Email: Priyanka.adusumilli@accenture.com

Priyank Patel
Accenture
1160 W. Swedesford Rd.
Berwyn / 19312
Email: priyank.h.patel@accenture.com