

## **Due (Data) Diligence: Study Data Review for Acquisitions**

Adam Young, ViroPharma Incorporated, Exton, PA  
Steve Kirby, ViroPharma Incorporated, Exton, PA

### **ABSTRACT/INTRODUCTION**

Drug companies are increasingly using acquisitions to leverage promising drug candidates. Ensuring that the price paid for a drug candidate is supported by clinical results naturally demands thorough review prior to purchase; but drilling down to investigate whether those results can be easily traced back to the clinical data can be neglected until close to the time of a regulatory submission.

As companies work to identify and purchase promising drugs, available study data should be carefully evaluated for consistency with regulatory submission requirements. Small deviations from submission requirements for study data can have a dramatic effect on timelines and cost. If those issues are identified early in the acquisitions process, mitigation plans can be put into place to minimize downstream impact and price can be adjusted to reflect the added time and effort needed to get the drug to market.

### **DATA STANDARDS AND THE FDA**

This paper focuses on the common situation where electronic submission of human clinical trials data to US regulatory authorities is part of the development plan for the acquired drug candidate. Here are a few representative quotes that highlight the importance that the FDA currently places on receiving standardized (CDISC) study data supported by complete documentation.

“Standardizing study data makes the data more useful. Data that are standardized are easier to understand, analyze, review, and synthesize in an integrated manner in a single study or multiple studies, thereby enabling more effective regulatory decisions.” [1]

“Data that are not standardized diminish the Agency’s ability to review the data efficiently, resulting in manual, labor-intensive processes and inherent inefficiencies in the review. They also limit the ability to automate some routine analyses.” [1]

“The Center for Drug Evaluation and Research (CDER) is strongly encouraging sponsors to submit data in standard form as a key part of its efforts to continue with advancement of review efficiency and quality. CDER has been collaborating with CDISC, a standards development organization (SDO), in the development of standards to represent study data submitted in support of regulatory applications.” [2]

“A critical component of data submission is the define file...An insufficiently documented define file is a common deficiency that reviewers have noted.” [2]

### **THE CHALLENGE**

Identifying and assessing study data issues that could cost money and cause delay at submission time is challenging. Forming a review team that has broad familiarity with data collection, programming, CDISC data standards and submission requirements takes care. Supporting that team with a defined process takes planning, and reviewers will be working in an unfamiliar environment. Any part of the study data and associated documentation that is not consistent with submission needs is of concern, and assessing the impact of any inconsistency could require regulatory input.

### **EXAMPLE DATA REVIEW FINDINGS**

Consider a few examples of what could be found during study data review. Perhaps study data documentation is limited to mapping specifications for SDTM and ADaM data; or what if it cannot be conclusively determined which

data are the final locked data used to generate the report; or maybe the analysis data for efficacy is only available in SAS® datasets containing variable names with lengths greater than eight characters; or it could be that important PK/PD data are available (and were analyzed) in legacy format while other study data are wonderfully consistent with CDISC standards. These may not be make-or-break issues, but the potential impact on submission timelines and the cost of even minor complications can be great.

Take for instance the last example. If it's decided to convert the PK/PD data to SDTM/ADaM format for submission: then documentation would also have to be created to support the freshly standardized data; and the mapping work would need to be consistent with the CDISC standards as applied to the rest of the data; and best-practice indicates that the report be recreated; and if inconsistencies are discovered between the reports, even more work would be needed to resolve them. However, if the legacy data are submitted as is, submission of data that are not in standardized format is more likely to lead to questions from reviewers, and answering those questions will be challenging as the format will not be familiar or supported by standard programs.

## **THE IMPORTANCE OF DATA STANDARDS**

Drug companies strive to create streamlined processes that will support their data from collection on through submission. However, when reviewing acquired data, what is ideal for one party can be perplexing to another. When issues arise that could delay submission, the challenge of evaluating unfamiliar data can become even greater. What may initially seem to be a minor hitch could eventually lead to a lot of time and money spent to fill in the holes.

Regulatory bodies are expanding the role of standardized study data in the submission process. In recent years, the FDA has emphasized that submission of standardized study data and documentation packages expedites review and helps get drugs to patients faster. The FDA has supported efficient use of standardized data and documentation by training staff in data standards, creating tools that leverage standardized data and by incorporating standardized data validation into the submission process.

To better support efficient review of their candidates, drug companies have worked to generate, validate and submit study data and documentation in standardized format. And just as evaluation of submission data and documentation is necessary to facilitate efficient scientific review; evaluation of the available data and documentation supporting a promising drug candidate is an important part of determining the value of a purchase.

In the “best-case” scenario, a potential purchaser’s evaluation of available data and documentation would show, for each study, that available data and documentation was complete, verifiably compliant with CDISC standards and comprehensive enough to fully support downstream regulatory use.

Although the study data are clearly of central importance to any submission effort, we recognize that the data are only one part of the information needed to support a submission. Discussion of the potential cost and delay associated with other types of submission information that needs to follow defined conventions is beyond the scope of this paper. However, just as with complications in submission data availability, organization or format, other content that is not easily available or ready to submit will impact timelines and submission cost.

## **EVALUATION OF STUDY DATA AND DOCUMENTATION**

Study data evaluation (including associated documentation) as discussed in this paper is designed to investigate the potential impact of any deviations from the “best-case” scenario on downstream use. Any final decisions about how to approach submission of data that deviates from current best submission practices would be done in consultation with the reviewing body.

Evaluation of how well available study data mesh with the “best-case” scenario is naturally done on a case by case basis. Although deviations from best-case submission requirements occur on a continuum; the possible types of deviations can be usefully grouped into categories. As a purchase would typically cover data from a number of studies, it is reasonable to evaluate the content on a study by study basis when categorizing deviations.

While CDISC data standards are rapidly becoming a necessary part of the US submission process, whether or not the data are available in CDISC format is by no means a bright-line test for how challenging and costly it could be to insure the data are ready to submit. Far more important in terms of cost and concern are whether the data can easily be shown to contain all final data used in the report and whether the data meets the minimum requirements for electronic submission. That said, evaluating data that follow CDISC standards can be done much more efficiently than a similar evaluation of legacy content, especially when the data are supported by a define.xml, as standard tools support validation and general knowledge of what CDISC data look like is common.

It is worth taking a brief, general look at some of the basic questions a data assessment would need to ask and answer.

- Are all data available (including PK, deviations, trial design, etc)?
- Are available data easy to locate and work with?
- Is it obvious which data are the final, locked data?
- Are all data available in a format consistent with basic electronic submission requirements?
- Are all TLGs generated based on the submission data?
- Are tabulations data verifiably consistent with SDTM?

A “no” answer to any of these questions indicates that challenges are ahead before the data can be submitted, and generating a list of specific questions to be answered could provide a useful framework for data evaluation. As these sample questions suggest, a solid understanding of the flow of clinical data from collection to report and of CDISC standards is needed to effectively evaluate how data and documentation conform to the “best-case” scenario.

Assessment of the potential impact of any deviations from the “best-case” scenario also needs to be informed by how the data have been used so far. If the data have been collected but only used for informal internal reports, the impact of the data not being ready to submit is limited as work is to be expected prior to submission in the ordinary course of events. Where formal reports intended for submission have been generated, any inconsistencies with best-case submission standards can have a great impact as any change to the data will require that related report content be re-verified or redone.

Using the “best-case” scenario (results directly supported by complete data and documentation that are consistent with CDISC standards and ready to submit) as a baseline, we will categorize deviations from that standard and discuss the potential impact of those categories on a purchasing company’s ability to submit the study data to a regulatory agency.

## **“BEST CASE” DATA AND DOCUMENTATION**

When study results are supported by complete CDISC (SDTM and ADaM) data and comprehensive documentation (define documents, blankcrf.pdf and data guides): 1) any review of the content by the purchasing company can be done efficiently, and 2) additions or updates will not be needed prior to submission. While the specific details of how the standards were implemented could have a minor impact on downstream use (such as with pooled analysis) when study data are supported by ready to submit CDISC data and documentation, the path to submission is short and safe.

## **DEVIATION (FROM “BEST-CASE”) CATEGORIES**

Deviations from the “best-case” scenario can be usefully categorized by the functional area of the deviation. Key functional categories are 1) data format, 2) data documentation and 3) missing data. As the purpose, requirements and downstream use of tabulations and analysis data differ, we discuss each separately within each category. At the margins, issues related to data format, data documentation and data that are missing or incomplete blend into each other and a problem in one category can often mean issues in another.

## **DEVIATION OVERVIEW BY CATEGORY**

### **DATA FORMAT ISSUES**

Data that are not verifiably consistent with CDISC (SDTM and ADaM) standards are more challenging to evaluate and use; and data that do not meet minimum standards for submission will need to be updated. When data are changed, associated report content often needs to be redone.

### ***Tabulations Data***

Some common issues with the format of tabulations data are:

- Data are similar to SDTM but have serious errors in validation
- Data are a mixture of legacy and SDTM
- Data are legacy only
- Data do not meet minimum standards for electronic submission

Tabulations data that are not (verifiably) consistent with SDTM standards can cause challenges at submission. If all the tabulations data needed for submission are not consistent with an accepted version of the SDTM standard, discussion with the reviewing authority to clarify the appropriate path forward is indicated. The agreed path forward could be to map the data into SDTM and regenerate the report, causing cost and delay. If it is acceptable to submit the data as is, any sponsor or regulatory review will likely take longer as it is harder to process information arranged in an unfamiliar format.

Having data that are not consistent with SDTM but are otherwise appropriate for submission is a small problem in comparison with data that do not meet minimum standards for electronic submission. Submission data sets must be in SAS® XPORT format with variable names limited to 8 and variable labels limited to 40. [3] Data that do not meet the minimum format requirements for submission means that the data will (at the very least) need to be reformatted and the content verified as consistent with the data prior to formatting.

### ***Analysis Data***

Some common issues with the format of analysis data are:

- ADSL not available when SDTM used for tabulations
- Data are a mixture of legacy and ADaM
- Data are legacy only
- Data that are not consistent with FDA recommendations
- Data do not meet not minimum standards for electronic submission

Data that are not (verifiably) consistent with ADaM standards can cause challenges at submission; but while ADSL is expected when SDTM is used for tabulations, ADaM standards do not need to be followed in all cases. If ADaM data are not submitted then the CDER study data specification document encourages sponsors to at least ensure their data meets a few basic requirements such as including decoded versions of coded variables and “core” variables. [3] Reformatting the data to meet the minimum standards often requires such substantial adjustments that the associated report will need to be redone and verified.

## **DATA DOCUMENTATION ISSUES**

Data documentation that is missing or not ready to submit complicates downstream work in a few related ways. Evaluating the data for any purpose naturally becomes more challenging when the supporting documentation is missing. Data documentation takes a substantial amount of time to complete, and that process is much more challenging when done months or years after the data were generated. The process of generating the documentation carries with it some risk that potential issues with data quality may surface, which could lead to larger scale rework and potential reassessment of study results.

### ***Tabulations Data***

Some common issues with tabulations data documentation are:

- Documentation (define, blankcrf.pdf) is present but not ready to submit
- Documentation is missing

A define document linked to an annotated CRF is needed for submission. Creating or updating this content can be time-consuming.

### ***Analysis Data***

Some common issues with analysis data documentation are:

- Define document present but not ready to submit
- Define document incomplete, inaccurate or missing
- Programs used to generate analysis data missing or incomplete

When the define document is incomplete or not available, determining exactly how the analysis data were derived requires an extended look into the programs used to generate the data. If the programs used to generate the analysis data are missing, it is (at best) challenging to establish exactly how the analysis data were generated. In that case, the data, the documentation and any related report content will likely need to be redone.

### **MISSING DATA ISSUES**

The impact on the purchasing company depends heavily on exactly what data are missing.

### ***Tabulations Data***

Some common issues with missing tabulations data are:

- Trial Design/Deviations data are missing from SDTM
- CRF data are not fully represented in electronic data
- Data not available in electronic format (most common with Non-CRF data such as PK)

The absence of some study level data (such as trial design data and deviations) from an otherwise competent SDTM submission package has limited impact as that content would typically be otherwise available as needed (but not as preferred) to support review. If tabulations data listed in the report and possibly used for analysis are not available in electronic format, further investigation into whether they can be located or regenerated is needed. Needing to generate electronic submission data from documents (such as CRFs or even listings) is at best time-consuming and problematic.

### ***Analysis Data***

Some common issues with missing analysis data are:

- Analysis data are not available for key analyses
- Analysis data are used to generate report tables but not available

It is expected that key analyses will be supported by analysis data. Otherwise analysis data are not necessary; but must be submitted if actually used to create a table. Any needed but missing analysis content would make it so table results cannot be traced to the source information, and supporting traceability can easily require that the analysis be redone and re-verified before submission is possible.

## **CONCLUSION**

Most drug companies are working to support and improve how they prepare standardized (CDISC) data and documentation so they can more efficiently collect, study and submit information about promising drug candidates to

regulatory authorities; and regulatory authorities are working to support and enhance how they evaluate and use standardized data so they can more efficiently assess submission information and more quickly determine if a drug should be approved. This focus on standardization of study data is making the drug discovery and approval process better and more efficient than ever. Years from now, it very well could be that whenever you acquire a promising drug candidate from another company you can be confident that the data and documentation will conform with CDISC standards and be ready to submit; but right now, there is a lot of variability in how data standards are applied; and digging into those details should be included as part of assessing a potential acquisition.

## REFERENCES

- [1] Guidance for Industry, Providing Regulatory Submissions in Electronic Format – Standardized Study Data (Draft, February 2012)]
- [2] CDER Common Data Standards Issues Document (Version 1.1/December 2011)
- [3] Study Data Specifications (v2.0, July, 2012)

## ACKNOWLEDGMENTS

The authors would like to thank ViroPharma Incorporated for backing this paper and for their ongoing support of data standards implementation. Thank you also to our excellent clinical programming team, with special thanks to Tammy Steele for her diligent review of this paper. Lastly we would like to extend our gratitude to the 2013 PharmaSUG conference committee for allowing us to share this paper.

## CONTACT INFORMATION

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

Name: Adam Young  
Enterprise: ViroPharma Incorporated  
Address: 730 Stockton Dr  
City, State ZIP: Exton, PA, 19341  
E-mail: [adam.young@viropharma.com](mailto:adam.young@viropharma.com)  
Web: [www.viropharma.com](http://www.viropharma.com)

Name: Steve Kirby  
Enterprise: ViroPharma Incorporated  
Address: 730 Stockton Dr  
City, State ZIP: Exton, PA, 19341  
E-mail: [steven.kirby@viropharma.com](mailto:steven.kirby@viropharma.com)  
Web: [www.viropharma.com](http://www.viropharma.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.