PharmaSUG 2017 - Paper SS10

The Do's/Don'ts, An SDTM Validation Perspective. The should/shouldn't when explaining issues in the SDRG (Study Data Reviewer's Guide)

Tom Guinter, Independent Consultant, Data Standards and eSubmission. Birdsboro, PA 19508

ABSTRACT

Sponsors collecting clinical trial data, and subsequently submitting that data to regulators, are expected to ensure:

1. The data is of sufficient quality and completeness to support evaluation of the clinical trial results relative to the study objectives,
2. The data format is compliant with the regulatory and industry standards accepted for eSubmission.

We typically consider both points together, often describing the combination as validation. But this really overloads the term validation, we should think of it separately as validation and (standards) compliance checking, and consider other issues such as perspective. Different individual roles using and/or preparing the data have very different perspectives on what the validation and compliance should provide.

This paper reviews some of the common issues sponsors face in the validation and compliance checking activities when preparing data for eSubmission to regulators.

INTRODUCTION

Many of us think of validation and compliance checking as a chore, something that's only necessary for the eSubmission. But it's really a tool to help us understand whether the quality of our data will support the study expectations, regulatory review and eSubmission expectations. We should look at it as our friend, helping us achieve our goals.

DEFINITIONS

Validation (Source: Merriam-Webster):

1 a: to make legally valid: ratify

b: to grant official sanction to by marking

c: to confirm the validity of (an election);
also: to declare (a person) elected

2 a: to support or corroborate on a sound or authoritative basis

b: to recognize, establish, or illustrate the worthiness or legitimacy of

Compliance (Source: Merriam-Webster):

1 a: the act or process of complying to a desire, demand, proposal, or regimen or to coercion

b: conformity in fulfilling official requirements

2 a disposition to yield to others

3 the ability of an object to yield elastically when a force is applied: flexibility

Perspective (Source: Merriam-Webster):

1 a: the technique or process of representing on a plane or curved surface the spatial relation of objects as they might appear to the eye

b: a picture in perspective

2 a: the interrelation in which a subject or its parts are mentally viewed

b: the capacity to view things in their true relations or relative importance

3 a: a visible scene

b: a mental view or prospect

4 the appearance to the eye of objects in respect to their relative distance and positions

STUDY DATA TECHNICAL CONFORMANCE GUIDE (SDTCG)

The FDA SDTCG (Study Data Technical Conformance Guide) is available through the FDA Data Standards Catalog and website. It's currently updated quarterly, first published in December, 2014 with the latest version 3.3 published March 2017.

The entire SDTCG should be required reading and regular reference for everyone involved in any aspect of Clinical Trial data, including collecting, standardizing, analyzing and/or eSubmission.

Section 8 is particularly important with regards to this papers content, it covers Study Data Validation. The first line of section 8.1 summarizes the role the agency feels validation plays quite nicely: "Study data validation helps to ensure that the study data are compliant, useful, and will support meaningful review and analysis." This aligns with the definitions provided above:

- Validation = "useful"
- Compliance = "compliant" (with standards)
- Perspective = "support meaningful review and analysis".

The latest version of the SDTCG (March 2017) defines three types of Data Validation rules:

1. SDOs (Standards Development Organizations) conformance rules (e.g., CDISC)
2. FDA eCTD Technical Rejection Criteria that assess compliance
3. FDA Business and Validator rules

While each of these types of validation rules plays a different role from the agency perspective, and it's helpful for us to understand each role, the complete set/combination of rules should be considered in their entirety.

The last line of the SDTCG Appendix titled "Data Standards and Interoperable Data Exchange" states: "In summary, the goal of standardizing study data is to **make the data more useful** and to support semantically interoperable data exchange between sponsors, applicants, and the FDA such that it is **commonly understood by all parties**." The validation and compliance rules and processes that expect us to explain any known issues in the SDRG supports these principles.

As mentioned previously the entire SDTCG should be required reading and regular reference for everyone involved in any aspect of Clinical Trial data, including collecting, standardizing and/or eSubmission. **At a minimum, reading section 8 should be considered a prerequisite for reading this paper.**

SO WHAT DOES IT ALL MEAN TO US?

You probably think of validation/Compliance checking this way:

Hopefully you will think of it this way by the end of this paper:

THE DO'S AND DON'TS

Don't consider the published rules and/or vendor tools as the be-all end-all.

The published rules and tool vendor implementations are just a part of the validation and compliance process. There are many things that can't be programmatically checked (such as whether you chose the correct domain in many cases), and there are many things that should be data quality 101 and standards 101 issues that we should know and understand regardless of whether they're explicitly stated in published rules. In-other-words, there's a lot that we should know and do as normal process, and the published rules and vendor toolsets are just tools to help us with part of what we are expected to do.

Do – Validate using the most recently published validation rules.

Don't – Rely on the validation rules when the study started.

And the obvious question is: Why can't I use the validation rules when the study started? Why would I be expected to consider validation rules created and/or published after my study start?

The answer is: Perspective.

- The agencies will use the most recently published rules when evaluating your eSubmission. This means if you use earlier rules sets that don't contain some current rules, when you explain your issues in the SDRG you won't explain some issues that the agency will identify.
- Hopefully your goal is to facilitate the review, not just check-off that you've completed a task.
- Providing a complete rationale why validation issues are identified is helpful to individual(s) receiving and consuming the data.

Yes, to some this seems burdensome. But think about perspective and what our goals are, and think about if you were the one receiving the data and trying to understand it. Just because an explicit rule didn't exist when the study started doesn't mean that it isn't an important issue to the reviewer/consumer now. And just because an explicit rule didn't exist previously doesn't mean that the underlying principle the rule now addresses wasn't an issue previously. It just may be that the rule wasn't explicit at the time.

Do – Provide rationale in the SDRG why issues are not resolved or not necessary to resolve.

Don't – Simply say "not sponsor practice."

Why – That pesky perspective again.

Hopefully your goal is to facilitate review. When the agencies run their validation checks, they'll see the issues identified. Providing an answer "not sponsor practice" only says, yes, you know it too. It doesn't provide any reason/rationale why it's not the sponsor practice.

But then, why should you be responsible to explain your sponsor companies practice? Because you're preparing an eSubmission on behalf of your sponsor company, and the reviewer/consumer doesn't know the rationale why your sponsor company decided this. Again, remember perspective, the reviewer/consumer only sees that it isn't done. A clear rational of why it isn't done will help the reviewer/consumer understand if it is or should be important to them.

Do – Remember, issues identified are not necessarily "errors" or "false positives."

Don't – Change the data simply to avoid a validation issue.

Why – Perspective (again).

- Some issues are informational and not necessarily errors. And while a number of the informational rules (such as adding a permissible variable) are currently changing in validation rules (being removed), they exist in many of the validation reports already done and there will always be some rules that are informational and/or data dependent. Remember, you're intimate with your data but the reviewer/consumer isn't. From the reviewer/consumer's perspective it's helpful to understand when the standards have been adapted or extended. In-other-words, it's helpful to know what's different from other eSubmissions they may have received, or will receive in the future.
- Although the answer to this question is not so straightforward all the time. If it's a data quality issue (such as start date is after end date), then yes, the data should be queried at the source if possible, practical, and risk-justified. But not of-course changed in the eSubmission data unless

it's updated through a documented, traceable query process back to the source collection, usually approved in the data collection DMP (DataManagement Plan).

- In the end, remember science and regulation should drive the data collection and quality requirements, the validation/compliance rules are a tool to help us access our standards compliance and some data quality issues. If there's a data quality issue that can't be corrected in the source, then it's helpful to the reviewer/consumer to understand why it couldn't be corrected, or wasn't deemed necessary or important enough to address. If it's a standards compliance issue, if you can't correct it, or again decide it's not important enough to address, then it's similarly helpful to explain why you decided this. Changing something to simply avoid the identified issue would almost be hiding the issue and could be considered "the tail wagging the dog.

THE DO'S AND DON'TS / SHOULD AND SHOULDN'TS – SOME SPECIFICS

FDAC340 (CT2001) - Value not found in <u>non-extensible</u> codelist.

Try to avoid this situation, but if you have it:

Do – Explain this in the SDRG

Don't – Create a user-defined codelist and associate that codelist with the variable. This would be the tail wagging the dog and essentially hiding the issue from the reviewer/consumer. You've chosen to use values that aren't expected in the standard, this should be explained in the SDRG.

FDAC208 (SD0080), FDAC050 (SD0082) – Date after last disposition date.

If the study is locked, Do investigate

If the study is not locked

If the subject completed the study, Do investigate

If the subject has not yet completed the study, Don't change the data

This would be the Tail Wagging the Dog

This is "Catch 22". The rule is designed based on completed studies, and at this time there's no reasonable way to update the rule for ongoing studies.

In either case, provide explanations in the SDRG. For ongoing studies, this may be as simple as "Ongoing study, some subjects have not completed the study yet."

FDAC197 (SD2236, SD2237) - ACTARM/CD does not equal ARM/CD.

Most likely this is correct data representation, but clearly a study quality issue when subjects are not treated according to the plan. From a perspective standpoint, hopefully it's clear why the reviewers/consumers would want to know when this occurs and why.

Do – Explain each of the subjects in the SDRG, but at a high level.

Don't – Say "the data are what the data are," this doesn't provide any information on why this happened. But of-course don't duplicate any explanations provided in the CSR (Clinical Study Report), but it's fine to reference them.

FDAC031 (SD1076) - Model permissible variable added.
FDAC341 (CT2002), FDAC344 (CT2005) - Value added to <u>extensible</u> codelist.

Both of these are examples of issues that are not necessarily problems, but more likely informational based on your perspective. It's possible they're errors if you didn't intend to add variables and/or extend the codelists and should be quickly investigated, but when it's intentional it's fine.

Do – If correct, explain this in the SDRG with a simple, probably standard, explanation. Something like: "Data was collected for this permissible variable" or "Additional values were necessary to properly represent the collected data."

Don't – Over-analyze this. So to answer the obvious question, yes, this is a case where a standard "we did it" answer/confirmation is probably fine, because there's no more information necessary.

Remember - Perspective

FDAC113 (SD0006) – No baseline result.
FDAC021 (SD1077) – FDA expected variable not found.
FDAC036 (SD1082) – Variable length is too long.
FDAC124 (SD1083) – Missing xxxDY variable.
FDAC022 (SD1097) – No Treatment Emergent info.

There's not really a lot to say on these variables. These are agency requested in one/more of the: SDTCG, FDA business rules and/or the standards themselves. So:

Do – follow these rules

Don't – not follow these rules.

Provide explanations in the SDRG of why you are not in compliance at your own risk.

Note: Some of these rules are evolving as at the time of this paper, and undoubtedly as new SDO and/or agency rules are published. One is the expectation for baseline result. There is a new flag variable proposed for the SDTM associated with SDTMIG v3.3 (not finalized/published yet) for Last

Observation Before First Dose Flag (--LOBXFL). But unfortunately it isn't clear yet whether this will affect the agencies expectations for baseline flags (--BLFL). Please look for the final SDTMIG v3.3 and/or the SDTCG to clarify these expectations.

SUMMARY

Think of validation as compliance (with standards), data quality and perspective (of reviewers/consumers).

Hopefully think of validation and reporting validation issues in the SDRG as your friend, not a chore. Remember, rules are defined to help us understand if we're meeting the compliance and data quality expectations, and to allow us to explain when our data isn't perfect and why.

Hopefully you'll think of it this way:

Please note the Do's/Don'ts provided here are the author's opinion only, and do not reflect any standards or regulatory agency official positions. Science and regulation should drive the data collection and reporting needs, as well as the standards representations and data quality expectations.

The specific rules presented in the Do's/Don't are selected hopefully as a representative sample of the different types of rules, to allow you to consider the recommendations across the entire rules sets.

Lastly, remember the SDTCG closing statement:

> "In summary, the goal of standardizing study data is to **make the data more useful** and to support semantically interoperable data exchange between sponsors, applicants, and the FDA such that it is **commonly understood by all parties**."

In this authors opinion the validation/compliance rules are designed to help us achieve this goal.

Thanks. 😊

REFERENCES

CDISC Study Data Tabulation Model (SDTM) v1.4 and Study Data Tabulation Model Implementation Guide (SDTMIG) v3.2. http://www.cdisc.org/sdtm

FDA STUDY DATA TECHNICAL CONFORMANCE GUIDE. Guidance for Industry Providing Regulatory Submissions in Electronic Format – Standardized Study Data. FDA CDER, CBER

CONTACT INFORMATION

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

Name: Thomas Guinter
Enterprise: Independent Consultant
Address: 243 Spruce Dr.
City, State ZIP: Birdsboro, PA 19508
Work Phone: 610-905-0598
Fax:
E-mail: guinter@ptd.net
Web:
Twitter: