PharmaSUG 2023 - Paper SS-045

Optimizing Efficiency and Improving Data Quality through Meaningful Custom Fix Tips and Explanations

Jennifer Manzi & Julie Ann Hood, Pinnacle 21

ABSTRACT

Prior to submitting study data to any regulatory agency, the standardized data is evaluated using at least one validation tool to ensure data conformance and quality. The issues in the resulting report are then triaged, first to determine the cause of the issue, then to decide on the best course of action to resolve or address each one. Knowing which issues should be researched and which can only be addressed through explanations in the Reviewer's Guide is crucial to efficiently utilize a validation report. Guidance provided by an organization on how to approach validation issues can help save time and ensure the issue is addressed correctly, resulting in higher quality and consistent data across studies. This paper will focus on steps to create meaningful Fix Tips that will enable users to quickly evaluate researching and resolving an issue. It will also include examples of when an explanation is the best option and the details needed to create comprehensive standardized explanations.

INTRODUCTION: IT'S OK. EVERYONE HAS ISSUES.

Issues happen. Not only in life but in standardized submission data, as well. In order to create a complete submission package, a comprehensive reviewer's guide is recommended to be included for clinical, nonclinical, and analysis data and contain "special considerations or directions or conformance issues that may facilitate an FDA reviewer's use of the submitted data and may help the reviewer understand the relationships between the study report and the data" or "context for analysis datasets and terminology…"[3] However, logic indicates that the fewer the issues, the less problematic the review will be once the submission gets into the reviewers' hands (and the shorter the reviewer's guides will need to be!).

It follows that the validation of standardized study data sets should occur as early and often as possible during the life of the study and possibly even before actual participant data is collected, to catch any possible problems with eCRF design, planned data collection, or derivation or mapping issues. After seeing several of the same issues appear in the validation report and looking into how to resolve or address the issue, many validators become accustomed to the message, descriptions, and associated follow-up that should take place. But clinical data standardization can be a challenging concept to master, especially for those who are just starting in the field.

There are different software packages on the market to assess the conformance of standardized data sets, most of which contain rules based on documents published by CDISC and regulatory agencies. However, the messages and descriptions output in the validation report can sometimes be difficult to interpret, even though validation of standardized clinical trial data sets is no longer quite such a novel concept. Fix Tips are meant to be a tool to help users determine why a validation issue is present by guiding them through the process of researching the root cause of the issue and choosing the best option to resolve the issue.

You may be thinking, "Why should I invest in creating Fix Tips if the message and description of the issue should be clear enough to understand what the problem is?" While it may be true that the message and description of validation issues have been refined over the years to help give users more insight into the problem, output from error messages of any kind can still be tricky to read and even more difficult to know how to fix. For example, we've probably all experienced the "Error 404" message that greets us when a website we're trying to get to can't be located. But this doesn't tell us why we can't get there (Is the URL correct? Is the server working?) or how we as users can fix it (or maybe can't fix it). Similarly, regardless of the tool being used to validate CDISC data, guidance on places to look for the cause of the issue and what steps to take to correct it will help cut down on the precious commodity of time.

HOW TO EAT AN ELEPHANT

Now convinced that creating Fix Tips is a good idea, deciding how to start cultivating them is the first step. The list of issues that could fire for standardized data may be in the hundreds or even thousands, depending on the content being validated and the standards that are being used, making it seem overwhelming to determine a place to start. However, most rules are applied by standard, so breaking the list down by which standard it applies to can be a good way to pare down the list. Then, a strategic way to tackle this list is to ensure that you have fix tips for issues that are most important to you and your users first. One approach is to determine the most frequently firing issues. To do this, you could use a Top Issues report filtered by standard if one is available in your validation platform or you could simply gather several validation reports based on the standard you're focusing on, merge the issues into one data set, and perform a frequency count to find the ones that occur most often. From there, you could further rank the resulting issues by how adversely they'll affect the review, possibly based on categories such as Severity or Impact if that's present for the issues. Another approach is to prioritize the issues that are most often asked about. If there is a central repository where questions are posed, you could check to see which issues people have the most questions about. This could even be as simple as looking through your inbox to search for issues that you've asked or received questions about if you're one of the SMEs who gets contacted when validation report questions arise. Adding fix tips for these specific issues can help ensure that this important information is distributed, and the knowledge is transferred to the intended target audience.

Once a list of more focused issues is created, the next step is to determine the best course of action for each one. The best way to do this is to get a good understanding of the rule that is causing the issue to fire by reading the message or description of the rule if one exists within your validation platform. Some reports will even link issues to the original rule or documentation published by CDISC or FDA that the rule is based on, and it's a good idea to check there as well to ensure the concepts are understood. You may even also be able to look at the coding to see the algorithm of the rule that gives the specifics like what variables or values the check is actually using to evaluate the data.

Then, once there's a good grasp of the reason for the rule along with an understanding of what the validator is using to check the data, a thought exercise can be performed. Try to think about the different ways the value could be populated to make the check fire. Of course, this might be easier if there are validation reports available to view that contain the issue and you can see which records were flagged for it. Could it be caused by something in the data itself that needs to be looked at or changed? Or a mapping issue that a programmer needs to adjust the code for? Or maybe something in the metadata is triggering the issue. Then think about what could be done in each case to resolve the issue. You might find that there are multiple possible actions that could be taken for the rule based on the root cause. Or it may be that the issue cannot be fixed, in which case your Fix Tip would simply indicate that an explanation needs to be added to the Reviewer's Guide to explain the data.

ANATOMY OF A FIX TIP

The next step is to actually write out your fix tip(s) for the issue. Length could be a limitation in some cases, and it's best to know this beforehand. Sometimes, it can be a challenge to express more detailed concepts succinctly, so be prepared that it may take a few iterations to condense the main points of your statements while still conveying the accurate and necessary information. However, if you've ever tried your hand at things like creating new values for Test Names in QS from questionnaire prompts or truncating inclusion/exclusion criteria to still contain the main ideas while remaining within the 200-character limit of IETEST, then you're bound to be a pro and sure to get the hang of it in no time! Also, make sure you're addressing your target audience. For example, if only programmers in your company ever review the validation report, adding programming code or macro references to the tip may be appropriate. However, if people in other roles outside of programming review the validation report and need to resolve issues, this may not be as helpful to them to recognize and understand what needs to be done to resolve the issue.

That said, it's possible to produce these recommendations with as little or as much granularity as required. For example, CT2002 is an issue that nearly every submission contains, and it can fire for all domains in all versions of SDTMIG, SENDIG, and ADaMIG. If overarching fix tips with the broadest reach that can be applied to almost any situation are needed for this issue, that can be done by simply reminding the user to ensure the correct term has been used as present in the CDISC Submission Values

in the CDISC Controlled Terminology or suggesting that they check to make sure no special characters or spaces are found in the value. For the same rule, if a Fix Tip is needed to apply to a specific standard, this could be achieved by mentioning the standard in the text and the information that needs to be relayed, such as, “This issue is common for ADaM data sets when a term is added to the DTYPE codelist”. Or Fix Tips for this issue could even be created to address value-level issues, such as noting that the terms “MULTIPLE” or “OTHER” are used in examples in the various SDTMIG versions for variables where the associated codelist does not contain those terms.

For cases where the Fix Tip is merely a prompt for the user to add an explanation to the Reviewer’s Guide for the issue, understanding the cause of the issue thoroughly can also help create a comprehensive explanation. Some general guidelines for developing complete and correct explanations are to provide as much detail as necessary, while also being concise, much like creating Fix Tips themselves! Reviewers need detailed information that explains the reasons why the issue is still present at the time of submission, why it could not be corrected, and what the effect is; not simply just text reiterating that the issue wasn’t corrected. Explanations such as “As per data”, “Data as collected”, or “Mapped as-is” are honestly not useful to a reviewer.[2] For more information on crafting ideal explanations, check out Kristin Kelly’s papers, “Best Practice for Explaining Validation Results in the Study Data Reviewer’s Guide” and the update to the original, “Explanations in the Study Data Reviewer’s Guide: How’s It Going?”.[2,3]

The last thing you may want to consider when creating Fix Tips is where to keep or maintain them. If your validation tool includes some type of repository where these can be stored, then that’s great! Your Fix Tips are sure to reach those who are most likely to benefit from them. However, if your tool does not contain an intrinsic solution, you can still create your Fix Tips in a searchable document and then store the file somewhere where it can be accessed by those who are running or receiving validation reports to process.

LET’S DO THIS!

SIMPLE EXAMPLE: SD1130

So, let’s put creating some Fix Tips into practice. In this scenario, one of the issues in our validation report is SD1130 (Inconsistent value for QNAM within QLABEL), which can be found in the FDA Validator Rules v1.6 and PMDA Validation Rules Version 4.0. Looking at the FDA Validator Rules, we can see that the Publisher ID contains “FDAB0009”.[3,7]

version 1.6, finalized December 2022

FDA Validator Rule ID	Publisher	Publisher ID	FDA Validator Rule Message	FDA Validator Rule Description
SD1130	FDA	FDAB009	Inconsistent value for QNAM within QLABEL	All values of Qualifier Variable Name (QNAM) should be the same for a given value of Qualifier Variable Label (QLABEL).

Figure 1. FDA Validator Rules v1.6

This lets us know that the reason for this rule can be found in the FDA Business Rules. Opening that document and searching for the issue shows the following:

Version 1.5 finalized June 2019

FDA Business Rule ID	FDA Business Rule
Clinical and Nonclinical	
FDAB009	All paired variables should have a one-to-one relationship. Examples include short name and name of test; parameter name and parameter code or number; variable name and variable label, etc.

Figure 2. FDA Business Rules v1.5

Applying this to SD1130, we can see that this rule was created to ensure that for every value of QNAM, there was only one value for its associated QLABEL. Using that information, we can create a Fix Tip like,

"In the suppqual data set, verify each QNAM only has one QLABEL associated with it". This will instruct the user where to look for the issue (i.e., In the supplemental qualifier data set, check QNAM and QLABEL values).

ADVANCED EXAMPLE: SD1144

A more advanced example would be a scenario in which there may be multiple ways to handle the issue. If we look at SD1144 (MHSTDTC date is after RFSTDTC) in the FDA Validator Rules, we can see that it is linked with Publisher ID CG0079.[3]

version 1.6, finalized December 2022

FDA Validator Rule ID	Publisher	Publisher ID	FDA Validator Rule Message	FDA Validator Rule Description
SD1144	CDISC	CG0079	MHSTDTC date is after RFSTDTC	The medical history dataset includes the subject's prior history at the start of the trial. Start Date/Time of Medical History Event (MHSTDTC) should be before Subject Reference Start Date/Time (RFSTDTC).

Figure 3. FDA Validator Rules v1.6

Since the Publisher ID begins with "CG", it can be found in the CDISC Conformance rules.

Rule ID	SDTMIG Version	Rule Version	Class	Domain	Variable	Condition	Rule	Document	Section	Item	Cited Guidance
CG0079	3.2	1	EVT	MH	MHSTDTC	MHSTDTC ^= null	MHSTDTC < DM.RFSTDTC	IG v3.2	IG v3.2[2.6]||IG v3.2[4.1.2.6]||IG v3.2[6.2]		IG v3.2[2.6][Section 2.6: Do not create separate domains based on time, rather represent both prior and current observations in a domain (e.g., CM for all non-study medications). Note that AE and MH are an exception to this best practice because of regulatory reporting needs.]||IG v3.2[4.1.2.6][Section 4.1.2.6, Item 2.B.3.A (Discussion of --CAT/--SCAT): Adverse Events (AE), Medical History (MH) and Clinical Events (CE), for example, are conceptually the same data, the only differences being when the event started relative to the study start and whether the event is considered a regulatory reportable adverse event in the study.]||IG v3.2[6.2][Domain Code Table: The medical history dataset includes the subject's prior history at the start of the trial.]||MH: The medical history dataset includes the subject's prior history at the start of the trial.]||Assumption 1:The Medical History dataset generally includes the subject's prior and concomitant conditions at the start of the trial.]
CG0079	3.3	1	EVT	MH	MHSTDTC	MHSTDTC ^= null	MHSTDTC < DM.RFSTDTC	IG v3.3	IG v3.3[2.6]||IG v3.3[4.2.6]||IG v3.3[6.2]		IG v3.3[2.6][Do not create separate domains based on time; rather, represent both prior and current observations in a domain (e.g., CM for all non-study medications). Note that AE and MH are an exception to this best practice because of regulatory reporting needs]||IG v3.3[4.2.6][Item 2.B.3.A (Discussion of --CAT/--SCAT): Adverse Events (AE), Medical History (MH), and Clinical Events (CE), for example, are conceptually the same data, the only differences being when the event started relative to the study start and whether the event is considered a regulatory reportable adverse event in the study.]||IG v3.3[6.2][Domain Code Table: The medical history dataset includes the subject's prior history at the start of the trial.]||MH: The medical history dataset includes the subject's prior history at the start of the trial.]
CG0079	3.4	1	EVT	MH	MHSTDTC	MHSTDTC ^= null	MHSTDTC < DM.RFSTDTC	IG v3.4	IG v3.3[2.6]||IG v3.3[4.2.6]||IG v3.3[6.2]		IG v3.3[2.6][Do not create separate domains based on time; rather, represent both prior and current observations in a domain (e.g., CM for all non-study medications). Note that AE and MH are an exception to this best practice because of regulatory reporting needs]||IG v3.3[4.2.6][Item 2.B.3.A (Discussion of --CAT/--SCAT): Adverse Events (AE), Medical History (MH), and Clinical Events (CE), for example, are conceptually the same data, the only differences being when the event started relative to the study start and whether the event is considered a regulatory reportable adverse event in the study.]||IG v3.3[6.2][Domain Code Table: The medical history dataset includes the subject's prior history at the start of the trial.]||MH: The medical history dataset includes the subject's prior history at the start of the trial.]

Figure 4. SDTM and SDTMIG Conformance Rules v2.0

The SDTM and SDTMIG Conformance Rules v2.0 will report which versions of SDTMIG the rule applies to and includes helpful information such as the section, item, and guidance text found in the related Implementation Guide that the rule was created for. In this example, we can see that events included in the AE, CE, and MH domains may be medical event-type data, however, the reason for these different domains is due to the date of the start of the event or due to which events the study considers as regulatory reportable adverse events. From this, we can see that this rule was created to help determine whether the event was mapped to the appropriate domain. Using this information, we can create Fix Tips to address different types of situations that could cause this issue and guide the users to check for specific problems. For example, "Verify the derivation for RFSTDTC is correct and programmed correctly. If RFSTDTC is derived correctly, verify the medical history start date for this record and the date used to populate RFSTDTC are correct. If the dates are correct, determine if this has been correctly reported as a medical history event instead of an adverse event. If the date values are incorrect and cannot be fixed, explain the issue in the Reviewer's Guide."

CUSTOM STANDARDS AND IMPLEMENTATION

To really gain the most benefit from standardization, many companies have found it worthwhile to create their own custom standards based on their interpretation of the CDSIC standards. Although the implementation guides provide generous and multi-faceted examples (we see you, TA and TE data sets!), they cannot possibly account for all the variations in study design, data collection, and standardization

across the wide variety of studies that are using CDISC standards. This can mean anything from developing glorified work instructions or company-wide guidance documentation with instructions on how to derive certain variables to creating and using a dedicated set of metadata to initiate data set creation to developing a basic library of forms using CDASH intended to be used in all clinical trials. Depending on the company's size, this can also be accomplished by one or two individuals up to several different departments weighing in across all changes to "The Standard". And guess what, not everyone will agree on the same interpretation, sometimes resulting in issues showing up in the validation report.

Even with standards, regulatory authorities can disagree on the actual rules used to assess standardized data sets' adherence to the guidance, not solely the degree to which the issue will affect the review. When comparing the validation rules published by FDA and PMDA for SDTMIG v3.3, we can see that there are several rules implemented by FDA not found in the PMDA listing and one rule unique to PMDA not included in the FDA listing.[3,7] Add to that, there are cases when the guidance found in the Implementation Guide and the guidance from the Technical Conformance Guide don't align. For example, for SDTMIG v3.2, SDTM/SDTMIG conformance rule CG0073 had been developed to ensure EPOCH was null for records in DS where DSCAT='PROTOCOL MILESTONE'.[7] However, the FDA TCG's requirement that "EPOCH should be included for clinical subject-level observation" has been widely interpreted to apply to all records in DS.[3] Similarly, for DM, three rules were developed with reliance on ACTARM being populated as 'Screen Failure', however in the FDA TCG, it's requested for ARM, ARMCD, ACTARM, and ACTARMCD to be left null.[3,7]

For these types of discrepancies, it is generally believed that the regulatory guidance should be implemented rather than strict CDISC conformance due to the reasoning that the regulatory agency has specific reasons for these requests and they're also the entity with the power to approve or disapprove the submission. However, if it becomes apparent that implementation will cause an issue to appear in the validation report, having Fix Tips to alert users to these special cases can help remind them and save time that may have been spent looking into why the issue occurred or updating programming.

Likewise, even when there aren't discrepancies between agency and CDISC guidance, sometimes interpretation of the implementation guides can lead to issues arising in the validation reports for studies that follow the company's custom standard. For example, one of the issues that has been hotly debated in the industry is how to map multiple enrollments. To handle this issue, some companies have decided to implement the method where SUBJID is added as a variable in the main domain rather than as a supplemental qualifier. When that is done, the issue SD0058 (Variable appears in dataset, but is not in SDTM model) will appear in the validation report.

version 1.6, finalized December 2022

FDA Validator Rule ID	Publisher	Publisher ID	FDA Validator Rule Message	FDA Validator Rule Description
SD0058	CDISC	CG0013, CG0542, CG0564, CG0565, CG0566, CG0567, 2, 10, 30, 31, 32, 268, 269, 270	Variable appears in dataset, but is not in SDTM model	Only variables listed in SDTM model should appear in a dataset. New sponsor defined variables must not be added, and existing variables must not be renamed or modified.

Figure 5. FDA Validator Rules v1.6

In the FDA Validator Rules, there are several Publisher IDs listed from both the CDISC SDTM and SEND Conformance Rules.2

Rule ID	Rule	Document	Section	Item	Cited Guidance
CG0013	Variable = Model List of Allowed Variables for Observation Class	IG v3.2|Model v1.4	IG v3.2[2.5]|Model v1.4[3.2.22]		IG v3.2[2.5][Sponsors may not add any other variables. . . . Standard variables must not be renamed or modified for novel usage. Using SDTM-specified standard variable names.]|Model v1.4[3.2.22][Each observation can be described by a series of named variables.]
CG0013	Variable = Model List of Allowed Variables for Observation Class	IG v3.3|Model v1.7	IG v3.3[2.5]|Model v1.7[3.2.2 2.1]		IG v3.3[2.5][Sponsors may not add any other variables. . . . Standard variables must not be renamed or modified for novel usage]|IG v3.3[2.5][[Using SDTM-specified standard variable names]|Model v1.7[3.2.2 2.1][Each observation can be described by a series of named variables. Domain-specific variables, a concept introduced in SDTM v1.5, are for use in a limited number of designated domains and will be identified in the appropriate implementation guide.]
CG0013	Variable = Model List of Allowed Variables for Observation Class	IG v3.4|Model v2.0	2.5|Model Concepts and Terms -- Variables|The General Observation Classes		IG v3.4[2.5][Sponsors may not add any variables other than those described in the preceding three bullets. . . . Standard variables must not be renamed or modified for novel usage.]|Model 2.0[Model Concepts and Terms --Variables][Each observation consists of a series of named variables.]|Model 2.0[The General Observation Classes]Domain-specific variables, a concept introduced in SDTM v1.5, are for use in a limited number of designated domains and will be identified in the appropriate implementation guide.]

Figure 6. SDTM and SDTMIG Conformance Rules v2.0

Looking at the first rule in the SDTM and SDTMIG Conformance Rules, it's clear that this rule is based on the idea that variables should only be present in the domain that are listed in the SDTM model for that domain's class.[7] However, SUBJID is only a variable in the DM domain following the SDTM model.

Taking these documents into consideration, a Fix Tip could be added to direct users to simply explain this in the Reviewer's Guide, like "When this issue appears related to SUBJID, check mapping to determine if SUBJID is being populated for multiple enrollments. If yes, explain this in the reviewer's guide. If no, remove SUBJID from the domain."

This issue may also be triggered for standards that are following a Therapeutic Area User Guide for which non-standard variables have been proposed but have not yet been incorporated into a version of SDTM yet. For those cases, a similar Fix Tip may be helpful to let the user know that the issue is expected so that they don't spend time researching why the variable has been added to the domain and can proceed to add an explanation that details the rationale behind adding it to the main domain.

CONCLUSION

Creating Fix Tips can seem like a mammoth task when looking at all possible rules within a validation tool. However, breaking these into manageable chunks by standard, frequency within reports, importance, or repeatedly queried issues will help to focus the effort so that the Fix Tips most useful to the organization's users can be prioritized. Writing detailed, but concise Fix Tips with the end users in mind will provide guidance, transfer knowledge, and ultimately conserve efforts for those reviewing validation reports, no matter what tool is used.

REFERENCES

[1] Food and Drug Administration. 2022, December. "FDA Business Rules v1.5 June 2019." Accessed March 23, 2023. https://www.fda.gov/media/116935/download

[2] Food and Drug Administration. 2022, December. "FDA Validator Rules v1.6 December 2022." Accessed March 23, 2023. https://www.fda.gov/media/103587/download

[3] Food and Drug Administration, Center for Drug Evaluation and Research (CDER), & Center for Biologics Evaluation and Research (CBER). 2023, March. "Study data technical conformance guide v5.0." Accessed March 23, 2023. https://www.fda.gov/media/153632/download

[4] Kelly, Kristin. 2018. "Best Practice for Explaining Validation Results in the Study Data Reviewer's Guide". PharmaSUG 2018, Phuse US Connect 2018, and Phuse EU 2018. https://www.pharmasug.org/proceedings/2018/SS/PharmaSUG-2018-SS13.pdf

[5] Kelly, Kristin. 2023. "Explanations in the Study Data Reviewer's Guide: How's It Going?". PharmaSUG 2022. https://www.pharmasug.org/proceedings/2022/SS/PharmaSUG-2022-SS-152.pdf

[6] Pharmaceuticals and Medical Devices Agency. "PMDA Study Data Validation Rules Version 4.0 (2023-02-28)". 2023, February. Accessed March 23, 2023. https://www.pmda.go.jp/files/000250753.zip

[7] SDS Study Data Tabulation Model (SDTM) Conformance Rules Subteam. “SDTM and SDTMIG Conformance Rules v2.0”. 2021, July. Accessed March 23, 2023. https://www.cdisc.org/system/files/members/standard/foundational/SDTM_and_SDTMIG_Conformance_Rules_v2.0.xlsx

CONTACT INFORMATION

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

Jennifer Manzi
Pinnacle 21
jmanzi@pinnacle21.com
https://www.linkedin.com/in/jennifer-manzi-mcallister-9926357

Julie Ann Hood
Pinnacle 21
jhood@pinnacle21.com
https://www.linkedin.com/in/julie-ann-hood-40350525