ABSTRACT

In the FDA Study Data Technical Conformance Guide, first released in December, 2014, there are a couple of paragraphs on the topic of the two subject identifier variables defined in SDTM and some details of how FDA would like to see them implemented. Implicit in this section is the request that data from screen failure subjects be included in the study SDTM data. Although not a formal requirement (yet), FDA is very interested in receiving information about a study’s screen failures/rescreen attempts and has been quite persistent in continuing to ask to be provided that data by sponsors.

This paper discusses steps taken at different companies to meet that request. It may provide some guidance of its own if you are thinking about but have not yet taken any steps to capture and report information about your screen failure subjects.

INTRODUCTION

In the good old days, before CDISC standards were a serious consideration, screen failures were just counted. No further description of the reasons for failing screening. A listing would also be provided that showed the inclusion/exclusion criteria compliance. Re-screening was not mentioned anywhere, in a way that did not make it evident whether a subject was actually screened more than once.

As we will see below, the reality today is quite different.

TECHNICAL CONFORMANCE GUIDE (TCG)

While we are certain that almost all of the audience is well aware of what the Technical Conformance Guide is, we’ll provide a short introduction for those of you who don’t yet know what it is and still stopped by to read the paper.

Briefly, the document is provided as a FDA standard reference for Standardized Study Data for Submission Purposes, and is intended to be included as an insert in the Guidance for Industry Providing Regulatory Submissions in Electronic Format – Standardized Study Data document that is the real Industry Guidance. No meaningful explanation has been provided for the need for two documents; we think the idea may be that a Guide document must be much easier to review and update than a Guidance, so FDA puts the little details here and expects us to read and treasure both of them.

Please note the Guide is very clear that it contains non-binding recommendations, that the sponsor is not required to follow any of the content, but is supposed to provide something similar or something that still meets the standards, currently only one variable in the TS domain for SDTM. Does that mean you can
ignore the contents of this document? Yes, but at your own risk. Personally, we do not feel that we have had long and mostly pleasant careers in this industry by ignoring suggestions from FDA, so we tend to err on the side of accommodating them whenever possible.

**SUBJECT IDENTIFIER GUIDANCE IN THE TCG**

In the FDA Study Data Technical Conformance Guide, Technical Specifications Document, in the section titled 4.1.1.2 General SDTM Considerations, version 3.3, there exist three paragraphs that stress the roles of the two subject identifier variables that are part of the SDTM model, namely SUBJID and USUBJID. We have reproduced those paragraphs verbatim below, copied from page 9:

Subject Identifier (SUBJID)
The SUBJID is an ID of the entity (i.e., person) that participates in a trial. If the same subject is screened more than once in a trial, then the subject’s SUBJID should be different.

Unique Subject Identifier (USUBJID)
Each individual subject should be assigned a single unique identifier across the entire application. This is in addition to the subject ID (SUBJID) used to identify subjects in each study and its corresponding study report. An individual subject should have the exact same unique identifier across all datasets, including between SDTM and ADaM datasets. Subjects that participate in more than one study should maintain the same USUBJID across all studies. It is important to follow this convention to enable pooling of a single subject’s data across studies (e.g., a randomized control trial and an extension study).

Sponsors should not add leading or trailing spaces to the USUBJID variable in any dataset. For example, applications have been previously submitted in which the USUBJID variable for each individual subject appeared to be the same across datasets; however, in certain datasets, the actual entry had leading zeros added, or zeros added elsewhere in the entry. This does not allow for machine-readable matching of individual subject data across all datasets. Improper implementation of the USUBJID variable is a common error with applications and often requires sponsors to re-submit their data. [End of copied material]

Did you see anything there that was new or surprising? Certainly not the last paragraph; sponsors have had problems doing anything in a standard fashion for decades now. However, please examine the second sentence of the second paragraph, where it says that SUBJID should be used in the study and its report, not USUBJID. This implies a deprecation of the role of USUBJID in a clinical trial, because USUBJID is no longer to be used as the subject Identifier in a Clinical Study Report (CSR). This marks a significant change from previous patterns of use. Finally, note the last sentence of the first paragraph, where multiple SUBJID values are expected if the subject screens more than once in the trial. That suggests that FDA wants to see screen failure data in SDTM, otherwise why would you need multiple SUBJID values for the same subject? While not exactly forbidden, the likelihood of a sponsor allowing the same person to screen and be treated more than once in a trial is remote.

**NO BIG DEAL, RIGHT?**

Well, like so many things, it depends. Changing the role of USUBJID from a key variable to what is essentially an aid to pooling data will certainly affect a lot of sponsors and vendors, especially those who have invested time and money in standard system code that has USUBJID references hard coded into it. While it is true that USUBJID could remain as a key variable; multiple SUBJID values should always map to a single USUBJID value, so you can probably get by making the necessary change by adding SUBJID into the key structure after USUBJID. Be aware though, depending on how you assign USUBJID values, there will probably not be a strict ordering by SUBJID, making it hard to find subjects by browsing within the domains. In that case, you will find it infinitely preferable if you replace USUBJID in the key structure with SUBJID for individual trials and use USUBJID as a key only in those pooled reports.
The obvious thing that will need to change is that SUBJID will have to be added to every domain that contains subject level data where it doesn’t already exist. Note that is every domain containing subject level data except DM, so right about now, the idea that this isn’t going to be a simple and straightforward change might be occurring to you.

The short answer is that SDTM isn’t currently well engineered to handle data for subjects who were never treated. For example, relative study days are another item requested by FDA and are certainly part of the SDTM variables that can be added to every applicable domain (unlike SUBJID). However, recall the SDTM definition of a study day, that it is the difference between the date of interest and RFSTDTC, which is not going to ever have a value for a screen failure subject if you consider a screen failure as not actually entering the trial (where RFSTDTC is date of study enrollment, not the ICF signing date or some other date prior to the date of study enrollment). This may not seem like a big deal, and it may not be depending on the length of the screening period in your trials, but if you have a complicated, multiple day screening period, you and probably your agency reviewer will miss the relative days when they have to verify that certain steps were taken in the correct sequence and the interval between the steps is as it should be. It should also be mentioned that one of us has had experience with a very helpful CRO that managed to supply study dates for screen failures by ignoring the definition of a study day and making up a replacement value for RFSTDTC for the screen failure subjects. While well intentioned, this violates so many basic principles of SDTM that we hope it does not need further explanation to our readers.

Also, we do note that FDA is pretty clear that they expect a deviation from the SDTMIG, in that FDA requests

Screen failures, when provided, should be included as a record in DM with the ARM field left blank.

on page 11 of the TCG, as the second line under DM Do main (Demographics). While they strictly mention only the ARM variable, we assume they meant the other three variables (ARMCD, ACTARM and ACTARMCD) as well. This is a pretty big violation of the SDTM standard, as these four variables are required, which means they cannot be left blank. FDA does provide an explanatory footnote:

Although this convention is inconsistent with the SDTMIG, FDA recommends its use so that “Screen Failure” is not specified as a treatment arm.

We think we can take this to mean that FDA doesn’t want any other text to describe these subjects, so “Unrandomized” or “Not Treated” aren’t going to be welcome in the ARM variable either. We’ll revisit the topic later in the paper.

<table>
<thead>
<tr>
<th>Row</th>
<th>STUDYID</th>
<th>SUBJID</th>
<th>SITEID</th>
<th>USUBJID</th>
<th>ARM (SDTM)</th>
<th>ARM (FDA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABC</td>
<td>1001</td>
<td>001</td>
<td>ABC-001-1001</td>
<td>SCRNFIL</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ABC</td>
<td>1002</td>
<td>001</td>
<td>ABC-001-1002</td>
<td>PLACEBO</td>
<td>PLACEBO</td>
</tr>
<tr>
<td>3</td>
<td>ABC</td>
<td>1003</td>
<td>001</td>
<td>ABC-001-1001</td>
<td>ACTIVE</td>
<td>ACTIVE</td>
</tr>
<tr>
<td>4</td>
<td>ABC</td>
<td>1004</td>
<td>002</td>
<td>ABC-002-1003</td>
<td>ACTIVE</td>
<td>ACTIVE</td>
</tr>
<tr>
<td>5</td>
<td>ABC</td>
<td>1005</td>
<td>003</td>
<td>ABC-003-1002</td>
<td>PLACEBO</td>
<td>PLACEBO</td>
</tr>
</tbody>
</table>

Figure 2. ARM according to CDISC vs FDA expectation for Screen Failures

**WHO CARES ABOUT SCREEN FAILURES?**

A better question is why does FDA care about screen failures? Traditionally, screen failures have been reported in the CSR in an aggregated fashion, with little information other than number and reasons, possibly some demographics. However, it is reasonably clear that FDA would like the subject data that underlies the aggregated information. Data that is more granular allows a reviewer to assess study conduct, one thing that some statisticians fail to grasp as important, although it is the foundation that allows them to do the fun parts of their job, like prepare analyses and publish results. There is a reason
why so much of the study protocol is on the subject of how to conduct the trial; we work in a regulated industry and are expected to follow the instructions in the protocol to ensure that the results are not biased or influenced inappropriately, which would mean that the results are tainted or not considered valid. By examining the study entry process, a reviewer can see whether the sponsor or one of the investigators have done something unacceptable in allowing subjects to enroll or receive a specific treatment in the trial, which could call into question the validity of any results. Investigators have been known to enroll subjects that do not meet the entry criteria and even steer them towards a specific treatment, playing assignment games to try to ensure the subject ended up in a specific arm. Random assignment of subjects to treatment arms is not a trivial thing to ignore from an analysis perspective, but non-statisticians can overlook the importance, especially if one arm holds out the promise of life and the other does not. We always need to remember that clinical trials affect peoples’ lives and not every participant has the same motivations as an ethical sponsor (which of course are the only ones we would ever work for).

There is a second group that has been showing more and more interest in screen failure results and that is the refereed journals that publish our results if we are lucky. They are doing the same thing as agency reviewers, looking for reasons why the results could be questioned. Their reputations are on the line if and when they publish spurious research, so they are other consumers of your research data that have a vested interest in assessing study conduct.

Another group that will care about screen failures is your study development group. It allows them to track study conduct for themselves in real time, so they don’t get blindsided by reports of inappropriate behavior at a stage where it is too late to do anything about it. While they will be marking and tracking SF data on their own during the trial conduct, having it in your SDTM means they can receive standard, updated reports/plots from you and will probably cause them to think kind thoughts of you in the future.

Finally, the last group that should care about screen failure data is you. We would think that the interest of the prior-identified three groups should be sufficient incentive for you to take interest, but note the proper inclusion of screen failure data is not a trivial problem. You will have to think and try things and assess the outcome and rethink the solution to implementing them at your company, including collaborating with other groups, consequently growing as an individual and becoming a better provider of solutions in your workplace.

**WHAT DATA SHOULD BE COLLECTED FROM SCREEN FAILURES?**

This is not that easily answered of a question, as there is not clear guidance provided by the regulatory agencies. We are certain that if you actually cornered an agency reviewer and asked for guidance, they would say something along the lines of “Well, whatever is appropriate for your trial”, i.e., you figure it out. The good news is you have someone else to appeal to, and that is your study development group, the people who are responsible for and are paying for the trial. They are in the business of determining what information they want on subjects (including screen failures) and control whether a site will be paid to collect results from subjects who have already been determined that they will not be enrolling in the trial.

At a minimum, demographics should be available, as well as results that support the entry decision that was made and any safety issues that appeared during the time between the ICF signature date and the time of exclusion. On the other hand, if you have a lengthy and/or complicated entry process, you may have large amounts of assessments that should be treated like enrolled subject data, with every reasonable effort made to clean and organize the data.

You should not expect that all of the information available for an enrolled subject will be collected for a screen failure. Assessments like labs, ECGs, etc., cost money and your study development group is going to be less than thrilled to hear that you think they should pay for something that doesn’t really advance the trial, so that you can collect all possible information for potential future use. Even verbally collected items like medical history have costs associated with them and a site may not be willing and/or able to help resolve issues for a screen failure subject, especially if the subject is not a regular site patient but has been recruited to the site only for the trial. In any event, the final decision of what is available is theirs, and you should make every reasonable effort to fit this data into your SDTM.
One of the big decisions that need to be made here is whether a screen failure subject has a study disposition record or records. The issue is whether a screen failure is considered a study participant or not, and this is a question of philosophy more than anything else. Some organizations will want to include the subject as a study participant, others will want to exclude them as they were never actually enrolled in the trial. This is a question that should be resolved by the study development group; it is their trial after all, all you will need to do is follow (and hopefully guide) their decision. We are going to proceed under the basis that your study development group has decided that screen failures are not enrolled in the trial, so they do not have study disposition records. This decision does have consequences other than missing study days; we will return to the subject later.

**RESCREEN SUBJECTS ARE NOT YOUR FRIEND**

You can divide your study population into three categories; study enrollees who did not fail screening, screen failures who never enrolled, and study enrollees who screen failed at least once before attempting to screen again (we call them rescreens). The first two groups will add minimal complexity to the trial, but you can count on the rescreens to amp up the complexity quotient.

In a trial that does not allow rescreening, a subject only has one chance to qualify for the trial (the simplest possible model), but one that limits the possibility of allowing a subject that had a minor deficiency to eventually enroll. We think it unlikely that your study development group will be in favor of restricting enrollment. Once you allow a subject to rescreen, you have to take steps to capture some data twice, and make provisions for identifying him as a subject who has multiple SUBJIDs in the trial, so that he is assigned the same USUBJID value for each screening effort. There can also be even more complexity; do you report adverse events that occurred during the first screening attempt in the medical history of the second or later attempts? There isn’t a hard and fast answer to that.

The best rescreen case is if you allow only one rescreen effort, so that you don’t end up with more than one set of previous screening values, but we had recent experience in a trial that had one subject rescreen twice and finally enter the trial, even though the protocol expressly forbid more than one rescreening effort. Let this serve as a warning not to make data collection and remapping design efforts that do not allow for flexibility here. If you don’t, you will end up needing it at a most unwelcome time. Rescreen failures are less likely to cause problems to appear; they will still be missing the values that identify the subject as someone who was enrolled in the trial.

We recommend that you read a paper authored for the 2016 PharmaSUG, titled *Prepare for Re-entry: Challenges and Solutions for Handling Re-screened Subjects in SDTM*. In this paper, the authors went into considerably more detail about some of the issues you will find when setting up to handle rescreen subjects, so we don’t have to repeat the same material (Thank you Charity and Paul!!!!). We do disagree with their proposed solution, but it seems that it was driven by their choice to preserve USUBJID as a key variable. We have identified this as one of the few places where adopting the FDA recommended SUBJID variable as the key subject identifier will simplify your work effort. With a unique SUBJID as the identifier in your data, there is no need to combine screening efforts and present data as if it was from one subject instead of several, so no need for a rescreen epoch.

This brings up another topic, how extensive does your effort to identify subjects who have previously screened/enrolled in one of your application’s trials need to be. The USUBJID variable is the variable supposed to uniquely identify a subject across an application, where application is defined as a complete regulatory submission, which may be only one trial, but often contains significantly more than one. As a sponsor, you should realize that there are advantages to making the scope of your USUBJID as broad as possible. However, the downside to this is that you will need to make a reasonable effort to identify unique subjects across all that breadth. When you are in a situation where subjects are intended to enroll in multiple studies, such as a safety rollover extension trial, it is not difficult to identify unique subjects, and this is the most likely case where a subject will enroll in multiple trials.
You should account for the circumstance where a subject does attempt to enroll in multiple trials over time, especially when the compound being tested has a lengthy development program. Some sponsors exclude subjects previously enrolled in a trial from future trials, thinking that a more diverse treatment population is more important than meeting study enrollment targets quickly. Also, patient privacy rules make it very difficult for the sponsor to check whether a subject enrolled in a trial several years ago if the subject does not volunteer the information. Sponsors do not get copies of the ICF, so they can’t do handwriting comparisons, much less check the names (and subject’s names will occasionally change over time). We have seen subjects change gender in the middle of a trial, race is really a 19th century artificial construct that is not as fixed as racists dream, and we can no longer rely on getting a birth date from subjects in some countries. The abilities of a sponsor to cross-check enrollees across studies is very limited. Often, the best that can be done is to report the known cases from extension trials and hope that subjects are honest in reporting whether they have taken your compound in the past, which would give you some place to start looking. A simple question during the screening process is going to identify those subjects who remember that they had already screened for a trial of your compound.

**SOME THOUGHTS ON USUBJID**

While your SUBJID will also be determined by the study development group (hopefully with your guidance to not make the variable too long!), they are not really concerned about the USUBJID value. You’ll want them to like what you intend to do, but they are not going to be users of the data except in multiple trial reports. You’ll need to review your screening policies to make sure that you can easily identify subjects that have previously enrolled in a trial, thereby knowing the number of subjects exposed to your compound.

As an aside, we have been truly astounded by the level of ignorance and illiteracy exhibited by some practitioners of SDTM, who will tell you that there is a standard for USUBJID that you must adhere to. If you think so, may we direct you to the end of the second paragraph on page 8 of Section 4 of the SDTM Implementation Guide, version 3.2 where it clearly states: “Many sponsors concatenate values for the Study, Site and Subject into USUBJID, but this is not a requirement. It is acceptable to use any format for USUBJID, as long as the values are unique across all subjects per FDA guidance.”

Also, if your scope spans multiple studies, you may want to remove any reference to the study number from USUBJID. Consider what happens to your sort order when you have subject(s) who were previously assigned a USUBJID value in another study, so their USUBJID value includes their prior study number. The sort-order no longer reliably implies anything about the order of subject enrollment.

The final suggestion is keep the variable short and simple. Having recently worked with a CRO whose expansion of USUBJID was out to 22 characters (with the site number included twice!!), you simply lose a lot of space on the printed page, especially when the total number of subjects enrolled in the application was less than 999. A four-character code would have sufficed and meant the listings would have been fewer and much more easily read.

<table>
<thead>
<tr>
<th>Row</th>
<th>Parent study</th>
<th>STUDYID</th>
<th>SUBJID</th>
<th>SITEID</th>
<th>USUBJID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABC</td>
<td>1001</td>
<td>001</td>
<td>ABC-001-1001</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ABC</td>
<td>1002</td>
<td>001</td>
<td>ABC-001-1002</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Row</th>
<th>Study MNK extension to ABC</th>
<th>STUDYID</th>
<th>SUBJID</th>
<th>SITEID</th>
<th>USUBJID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MNK</td>
<td>1001</td>
<td>001</td>
<td>MNK-001-1001</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>MNK</td>
<td>1002</td>
<td>002</td>
<td>MNK-002-1002</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>MNK</td>
<td>1003</td>
<td>003</td>
<td>ABC-001-1001</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. SUBJID vs USUBJID. Most of us doing currently
Good versus better SDTM: Including Screen Failure Data in the Study SDTM?, continued

<table>
<thead>
<tr>
<th>Row</th>
<th>Parent study</th>
<th></th>
<th></th>
<th>USUBJID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABC</td>
<td>1001</td>
<td>001</td>
<td>5137</td>
</tr>
<tr>
<td>2</td>
<td>ABC</td>
<td>1002</td>
<td>001</td>
<td>5802</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Row</th>
<th>Study MNK extension to ABC</th>
<th></th>
<th></th>
<th>USUBJID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MNK</td>
<td>1001</td>
<td>001</td>
<td>5566</td>
</tr>
<tr>
<td>2</td>
<td>MNK</td>
<td>1002</td>
<td>002</td>
<td>5700</td>
</tr>
<tr>
<td>3</td>
<td>MNK</td>
<td>1003</td>
<td>003</td>
<td>5137</td>
</tr>
</tbody>
</table>

Figure 4. Proposed USUBJID with first number (5) identifying the application.

ISSUES WITH THE NEW AND IMPROVED VERSION OF USUBJID

The previous examples point out some things to be aware of and to be ready to handle. One, unlike the case where SUBJID is embedded in USUBJID, you should not expect that SUBJID will look anything like part of USUBJID, and you will be much better off if you make the extra effort to make sure they can never be confused with each other. Two, you will have to add SUBJID to your normal sort keys, and we really recommend that you remove USUBJID, if only to force users to get used to the change and not attempt to sleaze through by keeping USUBJID in the key variables. The FDA request here will not be a trivial change for many users, you should be aware that asking your coworkers to stop using USUBJID and use only SUBJID in a CSR will meet some resistance (It is to be hoped that tar and feathers will not be part of the resistance effort). Three, the subject order when sorting by SUBJID will not necessarily resemble the subject order when sorting by USUBJID. This will allow you to identify those colleagues who are trying not to change by the amount of steam and frustration emitted from their office space when they try to merge together two data sets that should have the same keys, but one data set still has USUBJID in the sort keys.

DATA VALIDATION ISSUES

We thought it appropriate to share some of the findings when you use the Pinnacle Validator tool to check your compliance to the SDTM rules. Briefly, it was not as positive of a result as one would have liked.

First, don’t waste your time evaluating a study unless you have prepared a define.xml for the study, and it has to be a version 2 one, where you can specify the sort keys for each domain. Otherwise, the tool assumes that your data is using USUBJID as a key, and does not think highly of your capabilities in data preparation.

Some messages were expected ("Variable appears in dataset, but is not in SDTM model" for every domain where SUBJID has been added), but some were not. The message "Variable is in wrong order within domain" appeared several times, with the root of the problem being that there was a variable (SUBJID) in the domain between USUBJID and ~SEQ, not that the variable sequence was out of order. Interestingly, this message did not appear for every domain where SUBJID was added, something that could use further investigation by the folks at Pinnacle when they have nothing else to do.

Rescreen subjects caused some interesting messages to appear. A message similar to "--DY variable value is not populated", showed up in many domains. This was due to the fact that the screen failure data records did not have the study days calculated because there was no RFSTDTC value on the DM record for the SUBJID, but the Validator tool found a RFSTDTC for the USUBJID identified subject and identified the values as missing because the values were assumed to be calculable. A message similar to "Duplicate value for --SEQ variable" was found in every domain where screen failure data appeared,
which in our case was only AE, IE and LB, because each subject had multiple –SEQ variables when evaluating them by USUBJID.

SOME PHILOSOPHICAL CONCERNS

As noted earlier, we expected to have problems implementing the FDA recommendation for treatment arm and code for screen failures and were not disappointed. It started with 4 separate messages (one each for ARM, ARMCD, ACTARM and ACTARMCD), possibly to reinforce what a bad idea it is to leave SDTM required variables as blank. Along with these expected messages, we got seven others, three for baseline values that were not found in domains where we had set baseline flags, no Disposition record, no Exposure records, and two messages about study dates not supplied (RFSTDTC/RFENDTC). All of these messages make logical sense because in our case the screen failures were never enrolled in the trial, so they don’t have any of this information. As a test, we modified the treatment variables to match the SDTM-compliant values (SCRNFAIL, Screen Failure) for the treatment arm variables and all eleven messages disappeared, making it quite evident that the Validator tool can handle screen failures properly when your study development group insists that screen failure subjects are not study participants, but only if you identify them according to SDTM guidance.

We suspect that these (and other) messages appear not due to the limitations of the Validator tool, but to a bigger issue. The last time we checked, the Validator tool was supposed to be used to identify problems in SDTM/ADaM compliance, not in recommendations established in the TCG, no matter how well intentioned. We find it somewhat worrisome that FDA thought it appropriate to issue a Guide that deliberately asks us not to conform to a published standard. There was a joke we told at an earlier company; “We loved standards so much, that was why we had so many different ones”. Do we really need two standards here, one for FDA and one for the rest of the world?

<table>
<thead>
<tr>
<th>Good (SDTM IG)</th>
<th>Better? (FDA TCG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCRNFIAL</td>
<td>Screen Failure</td>
</tr>
</tbody>
</table>

A cursory search of the review topics for the next SDTM Implementation Guide didn’t turn up any evidence that this change has ever been submitted to CDISC for review, but we might have missed it. We also feel it appropriate to reference an entry in the Study Data Validator Rules, version 1.0 where the FDA Validator Rule says for FDA Business Rule FDAB013 (line 79):

> All subjects should have at least one baseline observation (~BLFL = ‘Y’) in EG, LB, MB, MS, PC and VS domains, except for subjects who failed screening (ARMCD = 'SCRNFAIL') or were not fully assigned to an Arm (ARMCD = 'NOTASSGN') or were not treated (ACTARMCD = 'NOTTRT').

There must be some people involved in preparing the Data Validator Rules who didn’t read the Technical Conformance Guide as closely as they should have.

The end result here? Well, Tom Guinter’s excellent presentation on Data Validation (You may have some fun finding the paper on the CDISC website, but well worth tracking down) indicated that you should not be changing your study data to eliminate messages from the Validator tool, and the last time we checked, CDISC doesn’t review and approve NDA filings. So, we will be adding some explanation to the study cSDRG for the eleven messages, hopefully for the only time.
CONCLUSION

We hope that we have provided a road map to what you will need to do to properly include screen failure data in your study SDTM data. It isn’t as simple of a task as it first appears, but it isn’t that complicated if you plan your steps and take the appropriate actions, driven by what your study development team wants you to do.

REFERENCES

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


CONTACT INFORMATION

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

Name: Henry B. Winsor
Enterprise: Relypsa Inc
Address: 100 Cardinal Way
City, State ZIP: Redwood City, CA 94069
E-mail: hwinson@relypsa.com

Name: Mario Widel
Enterprise: Eli Lilly & Co
Address: Lilly Corporate Center
City, State ZIP: Indianapolis, IN, 46285
E-mail: mwide@lilly.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.