

Good Versus Better SDTM -- Date and Time Variables

Mario Widel, Roche Molecular Systems, Inc., Pleasanton, CA

Henry B. Winsor, WinsorWorks, Limited, San Mateo, CA

ABSTRACT

The SDTM rules are pretty clear about many things, but there are some areas where guidance is not provided and it is left to the user to implement how they see fit. While there are many implementations that do not violate the rules; there are also implementations that not only meet the rules but also offer extra utility to the end user. In this paper the authors will cover some examples involving the Date and Time variables; showing how making some effort to think about the large picture and design the unspecific parts of your SDTM data and metadata can make your data friendlier and more useful to end users.

INTRODUCTION

Why limit ourselves to do just a good (enough?) SDTM submission set when with a little more effort we can do a better one? The authors have seen numerous cases where a SDTM data set may be compliant with the rules, yet is not as useful to the end user as it could and should be. In our opinion, this is work that is penny-wise and dollar-foolish. An individual may save programming and documentation time, but if it makes the final reviewer spend his time to add things that you could have done, who benefits? Certainly not the agency reviewer and even more certainly not your company and/or client, who is hoping that the reviewer can get his job done with a minimum of fuss and without delays that slow the hoped-for approval process.

The authors present some real cases of data and metadata that they have seen, and offer some suggestions as to how the usefulness of the data could be improved.

CASE ONE -- --DY

The SDTM Implementation Guide lists the Core category for the various --DY and --STDY/--ENDY variables as permitted. While the logic for this seems pretty clear; the values used to derive the variable values may be missing, so they can't be required. However, some non-industrious persons have used permitted status as an excuse to omit the column entirely. In our opinion, these variables should be populated wherever the corresponding --DTC or --STDTC/--ENDTC variables exist within the data set. Relative study days give context to the reviewer; he can see immediately the relationship of a data record to the study start/end date without having to do the actual calculation or rely on a VISIT label to be accurate rather than reflecting what was intended to happen. Let's not forget that the calculation is pretty simple one, so there is no legitimate excuse for not including these columns in every data set where there are start/stop dates.

CASE TWO -- STANDARD --TPT, --TPTNUM AND --ELTM VARIABLES

There is no requirement that the time scales reported in individual module --TPT, --TPTNUM and --ELTM be consistent with the variables' use in other modules. This often reflects a case of less than ideal CRF design; a value of 10 for PPTPTNUM corresponds to a PPTPT value of "4 Hours", while the VSTPTNUM value of 4 may also correspond to VSTPT value of "4 Hours". This makes it difficult to easily combine data from multiple domains, and while the numbers are sufficiently useful for the needs of data entry, there is a better way when it comes time to include these numbers in your module.

The better solution is to review all cases where you will be using the --TPTNUM/--TPT variables in a particular study and choose the values for --TPTNUM so that they can be used in as many modules as possible and have some mnemonic meaning. For example, there is no requirement that the values used in --TPTNUM be positive contiguous integers, they only need to be numbers. You may wish to give the --TPTNUM variable a value of -45, which would represent the time point 45 minutes prior to your reference point.

Here is an example of suggested values:

```
LBTPTNUM, PPTPTNUM, VSTPTNUM = -45 => LBTPT, PPTPT, VSTPT = "Prior to Dosing"
PPTPTNUM                       = 100 => PPTPT = "One Hour Post Dosing"
PPTPTNUM, VSTPTNUM              = 200 => PPTPT, VSTPT = "Two Hours Post Dosing"
LBTPTNUM, PPTPTNUM, VSTPTNUM = 400 => LBTPT, PPTPT, VSTPT = "Four Hours Post Dosing"
PPTPTNUM, VSTPTNUM              = 600 => PPTPT, VSTPT = "Six Hours Post Dosing"
LBTPTNUM, PPTPTNUM, VSTPTNUM = 800 => LBTPT, PPTPT, VSTPT = "Eight Hours Post Dosing"
```

Note that while the numbers -45, 100, 200, 400, 600 and 800 are not contiguous, they are legal numbers for the --TPTNUM variables and allow the data to maintain a chronological sorted order, while additionally having the numbers be somewhat meaningful by themselves to a reviewer.

In a similar fashion, you should make the --ELTM variables as consistent as possible as well. Using the above example:

```
LBTPTNUM, PPTPTNUM, VSTPTNUM = -45 => LBELTM, PPELTM, VSTPT = "-PT00H45M"
PPTPTNUM                        = 100 => PPELTM = "PT01H00M"
PPTPTNUM, VSTPTNUM              = 200 => PPELTM, VSELTM = "PT02H00M"
LBTPTNUM, PPTPTNUM, VSTPTNUM = 400 => LBELTM, PPELTM, VSELTM = "PT04H00M"
PPTPTNUM, VSTPTNUM              = 600 => PPELTM, VSELTM = "PT06H00M"
LBTPTNUM, PPTPTNUM, VSTPTNUM = 800 => LBELTM, PPELTM, VSELTM = "PT08H00M"
```

Note the slight difference between our display of the ISO 8601 Duration formatted value and most examples you may have seen. We impose a bit of structure using zeroes as placeholders, so that the overall string value is as consistent as possible across all columns. This is allowed under the rules for ISO8601 durations and allows you to convey the spread of values that may exist in the study.

CASE THREE -- REPORTING ELAPSED TIME

Another twist on --TPTNUM values is use them to report the actual elapsed times instead of expected values. SDTM does not make it easy to report actual elapsed times, which we think should be part of the standard.

For example, if you look at the second table on page 45 of the Implementation Guide(v3.1.2), you will note the LBTPTNUM values of 1, 2, and 3 correspond to the actual elapsed times of 30 minutes post, 12 Hours and 5 minutes and 24 hours and 10 minutes. Why not use 30, 1205 and 2410 as the LBTPTNUM values, with corresponding LBTPT values of "30 Minutes Post-Dosing", "0-1205 Hours/Minutes", "1205-2410 Hours/Minutes"? The numeric values are easily calculated, while the generation for the text strings is not a terribly complicated macro.

Let us look at what happens when we use the Case 2 Examples in the situation where the Hour one draw was 5 minutes late while the Hour Two Collections were 5 minutes early. The values change as below:

```
LBTPTNUM, PPTPTNUM, VSTPTNUM = -45 => LBTPT, PPTPT, VSTPT = "Prior to Dosing"
PPTPTNUM                        = 105 => PPTPT = "One Hour, Five Minutes Post Dosing"
PPTPTNUM, VSTPTNUM              = 155 => PPTPT, VSTPT = "One Hour, 55 Minutes Post Dosing"
LBTPTNUM, PPTPTNUM, VSTPTNUM = 400 => LBTPT, PPTPT, VSTPT = "Four Hours Post Dosing"
PPTPTNUM, VSTPTNUM              = 600 => PPTPT, VSTPT = "Six Hours Post Dosing"
LBTPTNUM, PPTPTNUM, VSTPTNUM = 800 => LBTPT, PPTPT, VSTPT = "Eight Hours Post Dosing"
```

Note that this means you will not be able to combine data for a patient across modules unless the data is collected at the same time. In other words, you can't force a relationship if the relationship didn't actually exist.

Note also that the --ELTM variables do not change, even though two sets of --TPTNUM values do;

```
LBTPTNUM, PPTPTNUM, VSTPTNUM = -45 => LBELTM, PPELTM, VSTPT = "-PT00H45M"
PPTPTNUM                        = 105 => PPELTM = "PT01H00M"
PPTPTNUM, VSTPTNUM              = 155 => PPELTM, VSELTM = "PT02H00M"
LBTPTNUM, PPTPTNUM, VSTPTNUM = 400 => LBELTM, PPELTM, VSELTM = "PT04H00M"
PPTPTNUM, VSTPTNUM              = 600 => PPELTM, VSELTM = "PT06H00M"
LBTPTNUM, PPTPTNUM, VSTPTNUM = 800 => LBELTM, PPELTM, VSELTM = "PT08H00M"
```

Note that using this scheme will fail if there is a gross error in collection, such that the 1 hour collection was actually done at 2 hours post-dosing and that a measurement was also collected at that time for the 2 hour. The likelihood of this actually happening is almost nil, but the case of a missing measurement is not. We suggest that you would populate the --TPTNUM variable with the nominal expected time, the --STAT variable with "NOT DONE" and this convention should be clearly documented in the module specifications.

CASE FOUR -- EPOCH AND EPOCHDY

Last year, there was a paper that spent considerable time and arguing why the SE module needs to be completely populated as the second module as soon as clinical data is available (Winsor, 2010). One of the reasons for this is being able to populate the EPOCH variable in an events domain such as AE. Populating EPOCH allows us to communicate the treatment at the start of the AE without having to merge the data with another module. For example, knowing a particular event started in the Third Treatment EPOCH is more directly useful than reporting that it started on day 04 and let the reviewer make the necessary calculation.

At this time, we would also suggest that a new date and timing variable be added to the pantheon of date/time variables, the EPOCHDY. This variable would report the number of days since the start of the EPOCH. For example, the suppose EPOCHDY for the above event is 1, then it is clear to the reviewer when this event occurred in relationship to the start of the EPOCH instead of making him/her do the necessary calculations themselves. At present, this variable would need to be populated within AESUPP, but there is hope that it may be formally allowed in a future standard.

SUMMARY

While calculations keep most of us employed, we hope that you've seen the value in augmenting your SDTM data with some extras that will only help your reviewer and probably speed the review process.

REFERENCES

Study Data Tabulation Model (2008). Available at the CDISC website:
<http://www.cdisc.org/models/sdtm/v1.2/index.html>.

Study Data Tabulation Model Implementation Guide: Human Clinical Trials (2008). Available at the CDISC website: <http://www.cdisc.org/models/sdtm/v3.1.2/index.html>.

International Standard ISO 8601, Third Edition (2004). Available at the ISO Website:
http://www.iso.org/iso/catalogue_detail?csnumber=40874

Wood, Fred, Boyance, Adrienne (2010), *Checking for SDTM Compliance: The Need for Human Involvement* Pharmasug 2010

Winsor Henry B., Widel Mario (2010), *Creating SV and SE First* Pharmasug2010

CONTACT INFORMATION

Your comments and questions are valued and appreciated. Authors can be reached at:

Henry Winsor
WinsorWorks, Limited
San Mateo, CA 94403
Email: henry@winsorworks.com

Mario Widel
Roche Molecular Systems, Inc
Pleasanton, CA 94588 U.S.A.
Email: mario.widel@roche.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.