PharmaSUG 2023 - Paper DS-062

From Sea to Shining Sea- End to End discussion on PR and LB data

Abraham Yeh, Bayer Healthcare Pharmaceuticals, Whippany, NJ

ABSTRACT

From the clinical case report form (CRF), many big pharma and healthcare companies have standard sets that could fit various purposes, from the study design to achieving the study's goal.

The statistical analysis would be heavily dependent on these selections to fit the purpose of the CRF. On the other hand, we have standard tables that present the results using the data collected/analysis done and shared with the cross-functional study team.

First, two options about the prior radiotherapy pages with specific dosages of drugs. One is directly a text field containing all information that can create a listing for the team. The other is to have a separate field containing the real numerical value and the unit that could be transformed. In such ways, further analysis can be performed and evaluated.

Second, three types of Urine analysis were picked from the laboratory data. I will show different kinds of layouts on each and their implications. From one, frequency tables can be summarized. While on the other, further analysis using CTCAE version 5.0 could be utilized.

I will share live examples that the team encountered on multiple SDTM panels - to show how important it is to select the correct CRF as per the needs of the analysis. Furthermore, we could brainstorm on other cases.

"From Sea to Shining Sea" could be possible if we collect the best-fit selections tailored for the study/project. Then, "to shining sea" meant a shining future we understood fully when selecting the appropriate options.

INTRODUCTION

In the pharmaceutical industry, we have been talking about the standards for a long time. When we think about the standard, what comes to mind is one set of uniforms that would fit all. However, in real situations, people are not always wearing uniforms. When CDISC started over a decade ago, area experts discussed and reviewed these standards. There are many dialogues among almost all sentences and guidelines embedded in the standard. SDTM, SDTM Implementation guide, and ADaM models were created and suggested.

Luckily, there are various ways that the project would like to approach the analysis needs – some of them do not require heavy number crunching; hence a simple listing to store all the collected information might be sufficient for these conditions. While on the other, some 'deeper' statistics might be required to fulfill the purpose of the research. If everyone could think before they apply the standard on a cross-functional level, that would help the team. Ironically, and perhaps unfortunately, as people adopt the standard further, people tend to forget to think twice and carry the selected 'standard' CRF pages from the previous study and might easily miss the individual needs of the analysis for their project. Things are neither black nor white – we might have grey with different layers inside, which we should appreciate the diversity of selections we could choose.

One of the famous words we frequently hear in the data science community is E2E, denoted 'End to End'.

Let me share two real-life situations of cases in this paper from the CRF selection to the standard tables and/or listings that fit the analysis need on different scales. There are two different cases from the CRF design to their different spectrum of analysis potentials. We need to think thoroughly before the choice was made so we can reap the good results later with both efficiencies and fewer reworks as we can do things the first time right and build a sound data science society.

I would like to make this paper an introduction to the data science society so that we might surely have more cases to investigate and discuss – hoped this is a good start. From now on, I would encourage us to

choose the best-fitted one with the mind to think about what the results are like far in advance so we can proceed with the analysis with flying colors.

CASE 1: PRIOR RADIO-THERAPY (PR PANEL AND ITS IMPLEMENTATIONS)

If we pay some attention, we will find that there are several cases in the CDISC SDTM Implementation Guide where there are options to perform on the variables xxDOSE and xxDOSTXT (xx could be EC, EX, CM, PR, and SU).

					performed).	
PRDOSE	Dose	Num		Record Qualifier	Amount of PRTRT administered. Not populated when PRDOSTXT is populated.	Perm
PRDOSTXT	Dose Description	Char		Record Qualifier	Dosing information collected in text form. Examples: <1, 200-400. Not populated when PRDOSE is populated.	Perm

Figure 1. Caption for SDTM IG 3.2 on PRDOSE and PRDOSTXT

There are reasons for these options. If we think about the end-to-end process early when we start the CRF review on what to bring in – the team will be benefitted.

The necessary discussion points here for the team should be:

1. What is the objective of collecting such data?
2. Is the data we collected have the potential to be used for further exploratory analysis?
3. If we chose the one with just the xxDOSTXT – are we okay to live with it throughout the study and future submissions?

Catalog number 3.3 **Prior radiotherapies: Dose (cGy) – (full analysis set)**

	Treatment A N=xx (100%)			Treatment B N=xx (100%)			Total N=xx (100%)		
Intent	n	Mean	SD	n	Mean	SD	n	Mean	SD
Adjuvant	x	xx.x	xx.x	x	xx.x	xx.x	x	xx.x	xx.x
Curative	x	xx.x	xx.x	x	xx.x	xx.x	x	xx.x	xx.x
Palliative	x	xx.x	xx.x	x	xx.x	xx.x	x	xx.x	xx.x
Neo-adjuvant	x	xx.x	xx.x	x	xx.x	xx.x	x	xx.x	xx.x
Missing	x	xx.x	xx.x	x	xx.x	xx.x	x	xx.x	xx.x

Table-specific footnotes (to be included under all 'real' tables from this set in a CSR/integrated analysis):

- {Program, file pathway, date/time and other details of analysis}.

Catalog explanations and table options (will not be included under the 'real' tables in a CSR/integrated analysis):

- n includes only subjects with a record for dose

Figure 2. Sample table catalogs as suggested

As Figure 2 showed - if the team like this kind of table and needs to have all the collected data collected properly in advance. Then the recommendation would be to NOT use the xxDOSTXT option, even though we are aware that if you find numbers and perhaps units in the text field, it is possible for the SAS or R to retrieve these and did a lot of manual operations. If the xxDOSE as numerical values are collected in the first place, we can easily perform the analysis using the standard programs that many companies have to approach.

But there is one question here – then, under what kind of situations should we be selecting the xxDOSTXT variables?

Here are some possibilities:

1. From the objective of the study, such variables were of little interest to the team. Most of the time, some listings can be suggested to the TLF specs/catalogs, and the team agrees.
2. There are difficulties within the collection of the numbers as the numerical values from the sites – these can be complicated at times, especially in the medical history or prior radiotherapies as not every patient/subject was so clear about their real numbers. Under such situations, as the data analytics team member and the statisticians, we better inform the team in Figure 2. Analysis cannot be reached in advance. If the team agreed, we could make the collection simpler.
3. If we think about the analysis itself and its potential use downstream, collecting all kinds of 'not so useful information might not be meaningful. As it is said, 'Garbage in, garbage out!' then why should we start to collect these in the beginning?

There are more instances in that we could investigate and explore this. The critical information resides in an important question 'What is the purpose of collecting this information?' Think about the concept of 'end to end' again. If the team felt the tables were needed for the study, then another question should be asked: ' Can we grab such values in numerical formats easily?' If there are two positive responses of 'Yes' for the above questions, then the recommendation should be to use the xxDOSE options.

Even if the team just had the option to select the xxDOSTXT, we can still deliver a listing to show such collections, which is still helpful for the study team.

SUGGESTIONS FOR THE XXDOSE OR XXDOSETXT SELECTIONS

1. From the CRF design, a CAUTION must be exercised for any team that picks either xxDOSE or xxDOSTXT.

2. From the template on the TLF, especially if a frequency table is being presented or selected, please trace back to see if a proper selection was made prior to the start of the study. A recommendation on the template to alert the team to consider this well in advance is a good practice that might prevent difficulties in the future.

3. For the study team, please ensure these variables were picked up properly before selecting the frequency tables. It is not true that people think there is no effort for the STAT and that the analyst team could do anything without many troubles. In such cases, if the team chooses improper inputs, the resources that would be required would be significantly higher and, in most cases, – not necessarily if we did the right thing in the beginning, the first time.

CASE 2: LABORATORY (LB PANEL AND ITS IMPLEMENTATIONS)

Three types of CRF collect Urinalysis data (two in dipsticks and one using 24-hour assessment). With different intentions for each one – we better pick the right one per the analysis's needs.

We will start by looking at the CRF design.

The first one is shown in figure 3 below.

URINALYSIS - SINGLE DIPSTICK

Glucose, Urine	XLBNO4P ○ Normal ○ 1+ ○ 2+ ○ 3+ ○ 4+	LB.LBTESTCD(=GLUC) LB.LBTEST LB.LBEDCS(=UGLU) LB.LBORRES LB.LBSTRESC LB.LBCODLST(=XLBNO4P)	Mandatory
Protein, Urine	XLBNE3P ○ Negative ○ 1+ ○ 2+ ○ 3+	LB.LBTESTCD(=PROT) LB.LBTEST LB.LBEDCS(=UPRO) LB.LBORRES LB.LBSTRESC LB.LBCODLST(=XLBNE3P)	Mandatory

and Bilirubin, Erythrocytes, Hemoglobin, Nitrite, Ketones, Leukocytes, pH, Urobilinogen

Figure 3. Urinalysis - Single Dipstick Collections

The single dipstick method provided the different urinalysis levels needed for calculating the CTCAE grade.

Proteinuria	1+ proteinuria; urinary protein ≥ULN - <1.0 g/24 hrs	**Adult:** 2+ and 3+ proteinuria; urinary protein 1.0 - <3.5 g/24 hrs; **Pediatric:** Urine P/C (Protein/Creatinine) ratio 0.5 - 1.9	**Adult:** Urinary protein >=3.5 g/24 hrs; 4+ proteinuria; **Pediatric:** Urine P/C (Protein/Creatinine) ratio >1.9	-
Definition: A disorder characterized by laboratory test results that indicate the presence of excessive protein in the urine. It is predominantly albumin, but also globulin.				
Navigational Note: 24-hour urine collection takes precedence over dipstick				

Figure 4. Urinalysis – 24 hours just for reference only

From the clinical point of view, the appropriate method to measure proteinuria is not with a dipstick but rather with a full 24-hour urine sample protein dosing test. (Dipsticks are just a quick, semi-quantitative analysis used for a “ballpark” assessment on the spot, typically in ER or urgent care environment. When truly important from a clinical point of view, confirmation is almost always done on the 24-hour urine sample)

The proteinuria CTC code allows for grading based on both the dipstick as well as the full, 24-hour urine

sample dosing. It even says that the 24-hour method takes precedence over the dipstick. So, pragmatically speaking, we expect that when clinically indicated, investigators would resort to the 24-hour method or maybe even the protocols should require this method when important. In this paper, we are not discussing this method but thought it was useful information from the clinical perspective.

Qualitative Assessment

Glucose	XINTERP o Normal o Abnormal, clinically insignificant o Abnormal, clinically significant o Not applicable o Not assessable	LB.LBTESTCD(=GLUC) LB.LBTEST LB.LBEDCS(=UGLU) LB.LBORRES if result is 'not applicable': LB.LBSTAT(=NOT DONE) LB.LBREASND(=NOT APPLICABLE) else LB.LBCODLST(=XFIND) LB.LBSTRESC LB.LBCLSIG	Optional
Protein	XINTERP o Normal o Abnormal, clinically insignificant o Abnormal, clinically significant o Not applicable o Not assessable	LB.LBTESTCD(=PROT) LB.LBTEST LB.LBEDCS(=UPRO) LB.LBORRES if result is 'not applicable': LB.LBSTAT(=NOT DONE) LB.LBREASND(=NOT APPLICABLE) else LB.LBCODLST(=XFIND) LB.LBSTRESC LB.LBCLSIG	Optional

and Blood, Ketones, Bilirubin, Leukocytes, Urobilinogen, Nitrite, Specific Gravity, pH

Figure 5. Urinalysis Qualitative Assessment

Another option available for urinalysis assessment is called 'Urinalysis Qualitative Assessment' (please see Figure 5 as shown above). If this selection is used, the CTCAE grading cannot be graded, as the information needed to calculate the grade was not given if the team chose this approach. There are still many options available when it comes to the TLFs – primarily on summary statistics (e.g., calculating the frequencies and their related statistics, but not as the gradable CTCAE).

SUGGESTIONS FOR THE COLLECTION IN THE CASE REPORT FORM

The first question to the team should be as the following:

Do we want to get the CTCAE grading for our analysis?

Depending on the response from the team, there are at least three options that the team could choose:

1. If the CTCAE grading is needed and the budget for running the test is low, the first best pick would be to choose the Single Dipstick, and these collected results could be used for the grading without any problems.

2. If the CTCAE grading is required and there is a sufficient budget to run the laboratory test, then the suggested approach is to go for the 24-hour assessment to evaluate the best results.

3. The qualitative assessment can be used if there is no need for the CTCAE grading. Please note that the summary statistics for the value of abnormal and normal can be assessed with the limited ability to assess the results.

CONCLUSION

With the help of these two live and good examples we came across in the studies in the past few years. There are many more examples in other potential areas with these end-to-end processes. As a team together, we should bear in mind the following items:

1. No one is a lonely island – as the success of the study/compound is based on teams – not individuals. Everybody needs to be involved early so any future surprises or troubles can be reduced to a minimum. If all people can see the downstream effect and try to strive for the best by thinking in advance early enough, that would help the team's overall success.

2. Learning from experiences – As we are all growing and keep learning through our experiences, it is always a good practice to share what we have learned from Team A to Team B so all these experiences can be transferred to each other by taking the best advantages from each other. These two examples shown in the paper are real examples we came across during the study, and the teams encountered these challenges. We used a weekly meeting/ chat forum to share, and this can potentially be helpful to others and to rethink what is 'End-to-End' really means.

3. Do things the first time right – Once the team learns, we must bear these valuable experiences in our minds to avoid falling into the slippery slopes as we did not realize these experiences. One good way is to document these in the centralized 'experience sharing' dockets so people can easily refer to them. The other way is when we develop the CRF templates and the TLF catalogs; cross-referencing and linking are good ideas. With excellent and detailed documentation, plus the linkages to each other – it would be a great collaboration across teams.

During the conference, I would like to collect and learn from others on similar/different topics that people and teams are learning for the whole end-to-end process so we could benefit from each other using the best 'win-win' strategy.

ACKNOWLEDGMENTS

I want to thank Yueh-Ling Hsieh and Anupama Sheoran as the primary supporter of this paper. In addition, the Oncology Statistical Analytics Team 1 (OSAT1) members and the Oncology Data Analytics (ODA) members within Bayer helped the example cases through multiple discussions internally.

RECOMMENDED READING

- *CDISC SDTM version 3.2 Implementation Guide*
- *PHUSE 2016 paper*

  Grading Lab Toxicities using NCI- Common Terminology Criteria for Adverse Events (CTCAE) (lexjansen.com)

CONTACT INFORMATION

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

Abraham Yeh
Bayer Healthcare Pharma

100 Bayer Boulevard
Whippany, NJ, USA / 07981
abraham.yeh@bayer.com
www.bayer.com