PhUSE US Connect 2019

Paper DH08

Effective Presentations of CTCAE Graded Laboratory Data

Bill Coar, Axio Research, Seattle, WA, USA
Amber Randall, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

ABSTRACT

In clinical trials, laboratory data are assigned CTCAE grades to quantify the magnitude of abnormalities and to associate abnormal results with specific adverse events. Some parameters are associated with only a low or a high grade, while some can be graded bi-directionally and associated with multiple AEs.

Graded laboratory data are often presented in a shift table. Traditional shift tables are often not concise and can be difficult to interpret – especially with bi-directional grading.

Versions of CTCAE beyond 4.03 also complicate data presentations. Some parameters now are graded based on baseline values or the magnitude of the change from baseline. In these cases, summarizing a shift does not make sense.

We will consider presentations of CTCAE graded laboratory data and the analysis data sets that supports them, with a focus on clinical relevance as it relates to identifying meaningful treatment-emergent abnormalities.

INTRODUCTION

Summarization of laboratory data is often done via continuous statistics in summary table and/or figurative format. While this helps with general trends over time, it does not necessarily identify individual abnormalities over time. Reference ranges may be identified based on normal ranges, but it is known that normal ranges may be dependent on the laboratory and/or subject level characteristics (such as age and sex).

An alternative way to assess laboratory data is through categorical measures. A straight forward approach is to simply assign a value as normal or abnormal (low or high) based on individual specific normal ranges. If the appropriate data are collected, this can be further categorized as clinically significant, though we do not consider this in our discussion. Alternatively, observed values can be assigned a categorical value, or toxicity grade, based on industry accepted definitions. With this approach, the magnitude of the abnormality is captured in the measure rather than just information about normal/abnormal. We explore summarization using both approaches.

TREATMENT EMERGENT LABORATORY ABNORMALITIES

Often there is interest in identifying lab abnormalities that emerge after start of therapy. If a lab parameter was normal at baseline but became abnormal after receiving treatment, or if an existing abnormality worsened, then this might suggest a reaction from the study drug. Similarly, if an abnormality changed direction (from low to high or vice versa), that too would be treatment emergent and of particular interest. Identification of treatment emergent abnormalities and potentially clinically meaningful changes can be done via shift tables, though this may change with the implementation of CTC Version 5.0 as will be discussed later.

NORMAL/ABNORMAL

The binary classification of normal/abnormal is simple as long as normal ranges are provided. This at first may seem reasonable. However, experience shows that missing ranges are fairly common with local labs, or when working with interim data. A typical data structure for one subject may be:

Dataset 1: Low, Normal, and High

PARAMCD	AVISIT	AVAL	AVALCAT1	BASECAT1	SHIFT1
ALT	BASELINE	40	NORMAL	NORMAL	
ALT	WEEK4	80	HIGH	NORMAL	NORMAL-HIGH
ALT	WEEK8	50	NORMAL	NORMAL	NORMAL-NORMAL

This example shows a transient treatment emergent high ALT value where the increase resolved over time. Nevertheless, it is important to identify the original increase. Data in this form can be summarized in a shift table over time to display the frequencies/percentages of patients in categories defined by AVALCAT1 and BASECAT1. Information of shifts in this manner can be displayed in several ways, but a more typical approach is shown below.

Table 1: Shift between Low, Normal, and High

			Arm A			Arm B		
			Post Baseline			Post Baseline		
Parameter	Visit	Baseline	Low	Normal	High	Low	Normal	High
ALT	Week 4	Low	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)
		Normal	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)
		High	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)
	Week 8	Low	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)
		Normal	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)
		High	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)	xx(x.x%)
Etc								

When critiquing the approach in Table 1, we find some noteworthy attributes. First, if there are multiple visits, the table becomes lengthy, and likely consists of many cells of 0 (0.0%). The denominator is not shown. This may be reasonable for large studies, or final data where the denominator is relatively consistent. However, for interim reviews (such as a DMC), there is variable follow-up because of enrollment, so some patients have data at later visits while others may only have data for the early visits. Because of this, using a single denominator for all visits (ie, safety) inflates the denominator at visits with little data. This makes percentages underestimate the true incidence of the abnormality at that time point.

Treatment emergent issues are identifiable and shown in red. While this may be easy in a short example as displayed, emergent abnormalities can be overlooked if the table is lengthy. The abnormalities can be identified, but there is no information with respect to the magnitude. Furthermore, you can't identify abnormalities that worsened later in the study unless the change is from low-high or high-low because the shift will be simply state low-low or high-high. In this case, the table suggests no shift masking the fact that the original abnormality worsened. This is important in disease areas that might have abnormal labs as an underlying condition.

Updates may be considered to improve this presentation. If the data are captured, we could consider distinguishing between clinically significant and not clinically significant. However, this creates more rows and columns, and still lacks a quantitative way to assess magnitude of a shift. We could consider condensing the output to just display "worst" post-baseline value to identify potentially emergent concerns, and display the last post-baseline value to identify if it is transient. However, what does "worst" mean? Some lab parameters are bidirectional and can have concerning abnormalities if they are either low or high. To begin to address this, we consider CTCAE grading.

NCI-CTCAE

Veeragoni and Mathur [Phuse, 2016] provide an overview of CTCAE grading in oncology studies. There are things to note that we apply to this discussion. First, not all lab parameters are graded. Second, some grading requires clinical judgement. Our focus in not on the grading itself. Our focus is on utilization of the (existing) grade over time to identify potentially clinically meaningful changes in lab parameters.

The example data previously presented may now look like:

Dataset 2: CTCAE Grading

PARAMCD	AVISIT	AVAL	ATOXGR	BTOXGR	SHIFT1
ALT	BASELINE	40	GRADE 0	GRADE 0	
ALT	WEEK4	80	GRADE 1	GRADE 0	1 GRADE SHIFT
ALT	WEEK8	50	GRADE 0	GRADE 0	0 GRADE SHIFT

Note the assignment of GRADE 0 for analysis purposes. There is no such categorization as GRADE 0 according to CTCAE. However, it is common to assign GRADE 0 to values not meeting any of the criteria for GRADE 1 or higher. Often, these are values within a normal range. Assigning GRADE 0 for analysis purposes is convenient and allows us to easily identify values that worsened over time. It also allows us to easily determine the magnitude of the shift in terms of the number of grade levels increased (or decreased). A typical shift table based on CTCAE grades is shown below.

Table 2: Shift in CTCAE Grade

			Arm A					Arm A				
			Post Baseline					Post Baseline				
Parameter	Visit	Baseline	G0	G1	G2	G3	G4	G0	G1	G2	G3	G4
ALT	Week 4	Grade 0	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)
		Grade 1	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)
		Grade 2	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)
		Grade 3	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)
		Grade 4	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)
	Etc											

In reviewing this table, we again see some noteworthy things. First is that we needed a smaller font to fit the necessary information on a page. This immediately makes things harder to read. Furthermore, there are many more rows and columns (and pages) to sift through, many of which will likely be cells containing 0 (0.0%). The magnitude of the shift can be inferred by looking at the difference between the baseline grade and the post-baseline grade. As with Table 1, it is not clear if the denominator is (or should be) the consistent over time.

A major flaw to the above presentation has to do with parameters that receive bi-directional grading. Some parameters have abnormalities in a single direction (ie, abnormal low or abnormal high) that are of clinical interest. Other parameters (such as sodium, calcium, and glucose) have bi-directional grading meaning that abnormalities in either direction are interest.

When we receive data as in Dataset 2 for any of the parameters with bi-directional grading, we receive a single grade. However, additional work is usually required to determine if it is an abnormal low or abnormal high. Normal ranges and alert flags can be helpful to identify the directionality, but awareness of missing values in these data is critical for clear interpretation.

A straightforward update to Table 2 to account for bi-directional grading is to add information regarding the directionality.

Table 3: Addition of Directionality

				Arm A					Arm A				
				Post Baseline					Post Baseline				
Parameter	**Trend**	**Visit**	**Baseline**	G0	G1	G2	G3	G4	G0	G1	G2	G3	G4
Sodium	High	Week 4	Grade 0	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)
			Grade 1	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)
			Grade 2	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)
			Grade 3	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)
			Grade 4	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)
	Low	Week 4	Grade 0	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)
			Grade 1	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)	Xx (x.x%)
			Etc										

The major challenge with this presentation is the ever-increasing numbers of rows, columns, and pages to sift through. As before many of these values will be 0 (0.0%). Shifts of clinical interest can easily be overlooked because of this.

An additional challenge is presented for bi-directional parameters where a subject may go from abnormal low to abnormal high. Consider the example in Dataset 3.

Dataset 3: Bi-directional grading

PARAMCD	AVISIT	AVAL	ATOXGR	BTOXGR	SHIFT1
SODIUM	BASELINE	131	GRADE 1		
SODIUM	WEEK4	153	GRADE 2		
SODIUM	WEEK8	148	GRADE 1		

In this case, it is not clear how to define the baseline toxicity grade without considering other data manipulations to identify the TREND column in Table 3. While a primary reason for an analysis dataset is to aid in reporting, CDISC standards may also need to be considered.

Assume the normal range for sodium is 135-145. When summarizing shifts in the low direction, we see that the GRADE 1 low actually resolves post-baseline, so the post-baseline GRADE may be assigned GRADE 0. Similarly, when summarizing the high direction, the baseline value is not abnormal high. Thus, the baseline grade may be assigned a GRADE 0.

Options exist to condense Table 3 to a more manageable size. Rather than summarizing by timepoint, it may be reasonable to summarize a shift to worst post-baseline value, and shift to last-observed post-baseline value. When directionality is considered, a worst post-baseline value is easily obtained.

CTCAE EVENT TERM

We propose a different way to present shift information that accommodates bi-directional grading and allows for easy identification of clinically meaningful worsening of lab parameters. First, we recognize that grading associates a lab parameter to a particular adverse event. While most grading is associated with an observed value outside of a normal range, some criteria are based on a fixed unit and value. For example, Grade 1 Hypertriglyceridemia is defined by a value between 150 mg/dL - 300 mg/dL, not the upper limit of the normal range. Thus, there are cases where a value many not be deemed 'abnormal' based on a normal range, but may be graded suggesting an abnormality.

When parameters with abnormalities in a single direction are of interest, then grading is associated with a single adverse event term. For example, only high values of ALT are of clinical interest, an abnormal high ALT is associated with the event term Alanine aminotransferase increased. Parameters with bi-directional grading are associated with two event terms: one for abnormal high and one for abnormal low. An example here is sodium. Abnormal low values are defined as hyponatremia whereas abnormal high values are defined as hypernatremia. Both are of clinical interest. We focus on summarizing by CTCAE event term rather than by lab parameter. In this manner, we remove the presentation of GRADE 0 and categorize the event as Not Present, as suggested by not meeting any criteria for Grade 1 or higher.

We consider first presenting information in a more vertical manner. In seeking a more condensed output, we choose to initially present baseline and worst post-baseline grades, and the magnitude of the shift. Additionally, shifts indicative of improvement are not generally of concern, thus are omitted. If meaningful shifts are identified, then the table can be requested by visit to identify when the abnormalities are occurring, and if they are transient.

Table 4: Vertical Presentation

CTCAE Term	Visit		Arm A	Arm B
Hyponatremia	Baseline	N	Xx	xx
		Not Present	Xx (x.x%)	Xx (x.x%)
		Grade 1	Xx (x.x%)	Xx (x.x%)
		Grade 2	Xx (x.x%)	Xx (x.x%)
		Grade 3	Xx (x.x%)	Xx (x.x%)
		Grade 4	Xx (x.x%)	Xx (x.x%)
	Worst Post-Baseline	N	Xx	xx
		Not Present	Xx (x.x%)	Xx (x.x%)
		Grade 1	Xx (x.x%)	Xx (x.x%)
		Grade 2	Xx (x.x%)	Xx (x.x%)
		Grade 3	Xx (x.x%)	Xx (x.x%)
		Grade 4	Xx (x.x%)	Xx (x.x%)
	Shift to Worst-Post Baseline	N	Xx	xx
		0 Grade Shift	Xx (x.x%)	Xx (x.x%)
		1 Grade Shift	Xx (x.x%)	Xx (x.x%)
		2 Grade Shift	Xx (x.x%)	Xx (x.x%)
		3 Grade Shift	Xx (x.x%)	Xx (x.x%)
		4 Grade Shift	Xx (x.x%)	Xx (x.x%)

This combines both lab parameter and direction of abnormality that describes whether or not a condition is present, and if so, how severe is it. Grade 0 is replaced by “Not Present” because the condition itself is not present, which intuitively makes sense. We choose to only display shifts indicative of worsening. For the case where baseline sodium is grade 1 low with post-baseline being grade 2 high, we assess this using two different CTC terms. The patient has grade 2 hyponatremia at baseline. For post baseline, they do not have hyponatremia, so the event is labeled as “Not present”. Thus, their shift is a -1 grade shift, which is an improvement with respect to hyponatremia.

Similarly, the grade 2 high post baseline indicates grade 2 hypernatremia post baseline. At baseline, hypernatremia was “Not Present”. The patient is then summarized as having a 2 grade shift in hypernatremia (not present -> grade 2). The “cumulative magnitude” of going from grade 1 low to grade 3 high is still masked in this case. Shifts from hyponatremia to hypernatremia are not presented, though we do not consider it here.

To accommodate summarization by event term rather than lab parameter, structure of the analysis dataset needs to be considered. Historically the BDS structure allowed for a single toxicity grade (as well as baseline toxicity grade). Our initial proposal defines two sets of variables associated with grading, a decision that has been further established with recent guidance from the ADaM team regarding ADLB[2]. We further add information about the CTCAE event term, as shown below.

Dataset 4: Additional columns for bi-directional grading

PARAMCD	**AVISIT**	**AVAL**	**ATOXGRL**	**BTOXGRL**	**ATOXGRH**	**BTOXGRH**
SODIUM	BASELINE	131	GRADE 1	GRADE 1	GRADE 0	GRADE 0

SODIUM	WEEK4	153	GRADE 0	GRADE 1	GRADE 2	GRADE 0
SODIUM	WEEK8	148	GRADE 0	GRADE 1	GRADE 1	GRADE 0
SODIUM	WPB		GRADE 0	GRADE 1	GRADE 2	GRADE 0

SHIFTL	**SHIFTH**	**CTCAEL**	**CTCAEH**
		HYPONATREMIA	HYPERNATREMIA
-1 GRADE SHIFT	2 GRADE SHIFT	HYPONATREMIA	HYPERNATREMIA
-1 GRADE SHIFT	1 GRADE SHIFT	HYPONATREMIA	HYPERNATREMIA
-1 GRADE SHIFT	2 GRADE SHIFT	HYPONATREMIA	HYPERNATREMIA

Parameters that only are graded in the low (or high) direction only have CTCAEL (or CTCAEH) populated. We recognize there is an art to structuring analysis datasets. This worked for us, and is consistent with current recommendations. Alternatives surely exist.

MOVING FORWARD

We should point out that some parameters such as creatinine in CTCAE Version 4.03 may be assigned a GRADE 1 simply because of an increase from baseline. However, the grading criteria are not strictly based on a change from baseline, and the rules allow for assignment of a baseline grade.

CTCAE v4.0 Term	**Grade 1**	**Grade 2**	**Grade 3**	**Grade 4**
Creatinine increased	>1 - 1.5 x baseline; >ULN - 1.5 x ULN	>1.5 - 3.0 x baseline; >1.5 - 3.0 x ULN	>3.0 baseline; >3.0 - 6.0 x ULN	>6.0 x ULN

More complicated challenges exist as we consider implementing CTCAE Version 5. The analyses of grades using CTCAE version 5.0 changes for some lab parameters. Consider the lab parameter alanine aminotransferase (ALT). The CTCAE term of “Alanine aminotransferase increased” under Version 4.03 is defined by:

CTCAE v4.0 Term	**Grade 1**	**Grade 2**	**Grade 3**	**Grade 4**
Alanine aminotransferase increased	>ULN - 3.0 x ULN	>3.0 - 5.0 x ULN	>5.0 - 20.0 x ULN	>20.0 x ULN

whereas the definition in 5.0 is defined by:

CTCAE v5.0 Term	**Grade 1**	**Grade 2**	**Grade 3**	**Grade 4**
Alanine aminotransferase increased	>ULN - 3.0 x ULN if baseline was normal; 1.5 - 3.0 x baseline if baseline was abnormal	>3.0 - 5.0 x ULN if baseline was normal; >3.0 - 5.0 x baseline if baseline was abnormal	>5.0 - 20.0 x ULN if baseline was normal; >5.0 - 20.0 x baseline if baseline was abnormal	>20.0 x ULN if baseline was normal; >20.0 x baseline if baseline was abnormal

With careful review, we see that the updated grading criteria depend of the baseline value. Thus, it is not clear if or how grading would be assigned to baseline values. By one interpretation, any grade itself implies a concerning change from baseline, thus an emerging abnormality. Thus, a “shift” is inherently part of the grade, so shift tables would not make sense. Summarization of the grade alone identifies emerging abnormalities. This will surely change how the results are presented.

Perhaps an approach is to simply modify Table 4 to only display post-baseline grades for event terms where grading inherently depends on the worsening from baseline.

A word of caution: will people recognize and be aware of the changes from V4.03 to V5? In our opinion, there is a fundamental change in the meaning of the grade. Shifts no longer exist for some parameters, but do for others. The new

presentation is likely to be confusing and careful explanation should be provided.

CONCLUSION

We see from the above examples that shift tables can and are used to identify treatment emergent abnormalities. Shift tables is their historical structure have been and continue to be challenging to interpret, and concerning shifts are often buried in rows/columns that are not technically of clinical interest. We present an alternative that collapses output to only display a worst post-baseline value and only display shifts of clinical interest.

We also note that changes in the grading definitions with Version 5 deserve attention. Programmers, statisticians, and medical writers all need to understand the grading system to properly program, present, and interpret the results.

REFERENCES

[1] Veeragoni,S., Mathur, A. “Grading Lab Toxicities using NCI- Common Terminology Criteria for Adverse Events (CTCAE)” Phuse 2016
[2] Analysis Data Model Implementation Guide, Draft Version 1.2

ACKNOWLEDGMENTS

The authors would like to thank Emily Woolley, Marcia Craib and Paul Stutzman for their careful review and input in the topics presented in this paper. The authors would also like to acknowledge the many programmers publishing information on the internet, and apologizes to those not properly referenced above, citing the inherent nature of internet searching which makes it difficult to track so many individual ideas.

CONTACT INFORMATION

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

Bill Coar
Axio Research
2601 4th Ave, Ste 200
Seattle, WA, USA 98121
Work Phone: 970-219-8564
Email: williamc@axioresearch.com