

# What Types of Safety Outputs Does the DMC Receive?



David Kerr and Nand Kishore Rawat

**Abstract** This chapter goes into detail about the outputs that are both comprehensive and comprehensible for the DMC – this chapter focuses on safety. Details of the standard safety outputs are provided. These primarily are based on adverse event data, and a background is provided of how this adverse event data is captured and categorized. Detailed list of adverse outputs of most use to the DMC is given, with numerous examples of figures. Proposals for useful summarizations (including figures) from deaths, laboratory data, vital signs data, and other data sources are also provided.

**Keywords** Safety data · Adverse events · Serious adverse events · Adverse events of special interest · CTCAE · MedDRA SOC/PT · Deaths · Laboratory data · Liver function tests (LFTs) · Hy’s Law · Vital signs

All post-baseline information provided to the Data Monitoring Committee (DMC) by treatment could theoretically reflect safety. Efficacy outputs will be discussed later, and in theory, the efficacy results could reveal safety concerns in the form of “reverse efficacy” – where the endpoint is showing a harmful unexpected trend. And disposition outputs could also reveal safety concerns (e.g., discontinuation from treatment or need for more concomitant medication could indicate harmful trend). This section will focus on the more traditional measures of safety, however.

The most common way to evaluate safety is through outputs of adverse events (AEs). Most commonly AEs are of interest when they are treatment-emergent – occurring

---

All figures created by Bill Coar.

---

D. Kerr  
Seattle, WA, USA  
e-mail: [david.kerr@cytel.com](mailto:david.kerr@cytel.com)

N. K. Rawat (✉)  
Lantheus Medical Imaging (United States), King of Prussia, PA, USA

© The Author(s), under exclusive license to Springer Nature  
Switzerland AG 2023

N. K. Rawat, D. Kerr (eds.), *Data Monitoring Committees (DMCs)*,  
[https://doi.org/10.1007/978-3-031-28760-2\\_11](https://doi.org/10.1007/978-3-031-28760-2_11)

after the first dose or intervention and including only subjects who have had at least one dose or had the intervention. On occasion, the DMC may want to look at a listing of AEs that were pre-treatment – for example, if part of the screening prior to starting treatment is to wean off a medication or an invasive scan is required, then the DMC might want to be aware of AEs during screening as well as the treatment-emergent AEs. The AE monitoring period typically extends through a set period of time after last dose (e.g., 28 days after last dose – or some number of days that represents four half-lives of the medication). The protocol and the sponsor and SDAC should clearly communicate what period of time is covered by the AE surveillance. Studies that have multiple parts (double-blind followed by open-label extension) should especially be clear on which AEs are summarized in which set of outputs. In some studies, particularly open-label studies, there might be a different schedule of visits for subjects on different arms. The more opportunities there are to ask a subject about AEs, the more likely to have instances of recall bias and therefore nominally higher rates.

AEs typically are categorized as serious (an SAE) or not. There is a formal definition of seriousness:

- results in death,
- is life-threatening,
- requires inpatient hospitalization or results in prolongation of existing hospitalization,
- results in persistent or significant disability/incapacity,
- is a congenital anomaly/birth defect,
- is a medically important event or reaction.

A core focus of the DMC will be on the by-arm summary of treatment-emergent SAEs.

AEs are typically categorized by severity grade – for example, mild vs. moderate vs. severe vs. life-threatening vs. fatal. Note that a severe AE is not necessarily serious, and vice versa. Grading might also be on a 1 vs. 2 vs. 3 vs. 4 vs. 5 scale. One standard grading approach is Common Terminology Criteria for Adverse Events (CTCAE). A standard output for the DMC is AEs summarized by maximum grade, and AEs that are Grade 3/Severe or worse.

AEs are commonly coded to MedDRA (Medical Dictionary for Regulatory Activities). It would be challenging to simply list all of the verbatim terms sites enter for each AE. Instead, a coding process is implemented to translate each verbatim term to a term in MedDRA. There are variants, but summaries of AEs are typically done with MedDRA at two levels – the System Organ Class (SOC) level which has 27 levels and then within SOC at the Preferred Term (PT) level which has nearly 20,000 unique terms. Note that due to the real-time nature of the data, some AEs might not yet be through coding at the time of the data snapshot. These should still be included in outputs, perhaps by showing the verbatim term entered. The DMC should be aware that very similar events might be coded into different PTs. It will not be immediately obvious if PT lines of a table can be added, or if simply summing the lines would lead to excess due to double-counting of subjects who show

up on different lines due to having multiple events that were coded differently. For example, if two subjects show up in a table with PT of “Neutropenia” within the “Blood and lymphatic system disorders” SOC and two subjects show up in the same table with PT of “Neutrophil count decreased” within the “Investigations” SOC, it is impossible within just these results to determine if these two summaries represent two, three, or four unique subjects. The DMC can request the SDAC provide outputs that aggregate certain “constellations” of terms together. There are Standardized MedDRA Queries (SMQs) that do this for some standard groupings also. The study team may also have identified AEs of Special Interest (AEoSI). In such a case, the DMC outputs will include a summary of the PTs that are included in the set of AEoSI. One aspect that confuses DMC members who have not previously seen data summarized by MedDRA is how the condition under investigation is handled. They may suspect there is a problem when, in a Crohn’s Disease study as an example, some subjects show up with a PT of “Crohn’s Disease.” The DMC member will say that it should either be 100% (because all subjects had the condition at baseline) or 0% (since it is not treatment-emergent). The answer is that these are recorded as AEs if the condition worsens, for example here, if there is a flare in the Crohn’s Disease, it would be captured as an AE. In some protocols, a worsening of condition under investigation would not be captured as an AE but would be separately reported on the form that collects primary and secondary endpoints. The DMC should be informed on where and how AEs are collected as they relate to the clinical indication.

AEs can be categorized at the site for causality (e.g., possibly related, definitely related). It is suggested that the DMC generally ignore this categorization. The DMC will review the by-arm outputs. If there are more events on the active arm, then that type of event is likely causally related to the intervention. There is no need for the DMC to review or disagree with the investigator assessment of causality.

Information may be obtained if the AE resulted in interruption in treatment, change (reduction) in treatment, or permanent withdrawal from treatment. Note that some studies collect data if an AE led to discontinuation from the study – not just withdrawal from treatment. In most studies, this option should not exist. An AE certainly can motivate permanent withdrawal from treatment, but there’s no reason that an AE should impact whether the patient stays on study and has data assessments collected in the future.

A standard set of outputs from AE data is shown here:

- Overall Summary of Treatment-Emergent Adverse Events.
  - At least one AE.
  - At least one SAE.
  - At least one Grade 3/Severe or higher AE.
  - AE leading to withdrawal from study drug.
  - AE leading to death.
- Treatment-Emergent Adverse Events by MedDRA SOC/PT.
- Treatment-Emergent Adverse Events by Descending Frequency of MedDRA PT.
- Treatment-Emergent Serious Adverse Events by MedDRA SOC/PT.

- Treatment-Emergent Grade 3/Severe or Higher Adverse Events by MedDRA SOC/PT.
- Treatment-Emergent Adverse Events Leading to Withdrawal from Study Drug by MedDRA SOC/PT.
- Treatment-Emergent Adverse Events Leading to Death by MedDRA SOC/PT.
- Treatment-Emergent Adverse Events of Special Interest by MedDRA SOC/PT.
- Treatment-Emergent Adverse Events by Maximum Grade by MedDRA SOC/PT.

These outputs are typically presented at a subject level – a subject will show up in the numerator if they have at least one of the events of interest – whether that be just a single event, or 2 or 5 to 10 events. Percentages represent the percent of subjects who had at least one of the events of interest. This typically isn't a concern, but a DMC might request additional information that provides insight into the total number of events on each arm, not just the number of subjects who had at least one event.

The DMC should be aware if there is differential premature discontinuation of treatment, the average on-treatment time could be different which again would impact the observed rate of AEs – but not due to any true difference in the AE profile. Another fine-tuning of AE outputs is to summarize AEs per 100 patient-years. This analysis will adjust if the average time under AE surveillance is different between the study arms. These analyses could include a subject at most once in the numerator or could include a subject multiple times if the subject had multiple of the event.

Typically, inferential statistics (e.g., p-values, confidence intervals) are not included in AE summary tables. Some DMC members have requested these, but there are concerns of misinterpretation. An AE summary table might go on for 10 pages, for example, representing all 200 unique preferred terms that have occurred at least once. Including p-values for each of these lines could easily be misinterpreted. Due to multiple comparison, one might expect – by chance alone – 10 lines to have p-value < 0.05. Looking at a p-value < 0.05 might be used as a flagging mechanism, but DMC members could easily mistake these AEs as demonstrating conclusive proof of difference.

The only listing of AE data typically included is the listing of SAEs. All other information can be provided on an as-needed basis by the SDAC. There is minimal value in extensive listings of AEs. One feature appreciated by DMCs is to include a cumulative listing of SAEs, but to highlight (in bold, or a different color font) the incremental SAEs that are new compared to the listing generated at the previous DMC meeting.

Figures based on AE data are less common but should become more common. Some examples are shown below. They can be helpful to look at differences, but do not entirely replace careful review of summary tables. An imbalance of 4 vs. 0 in progressive multifocal leukoencephalopathy (PML), for example, could be a critical topic for the DMC but likely would not stand out in the following graphics.

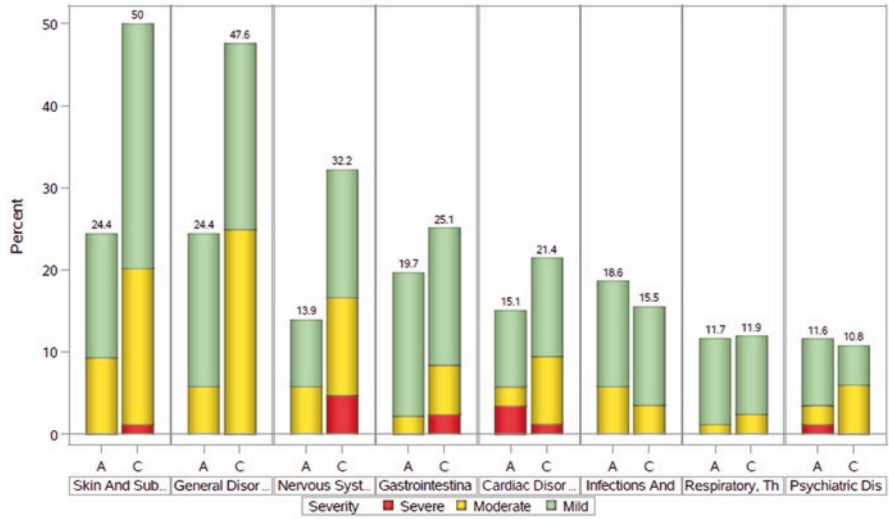


Fig. 1 Adverse events by System Organ Class

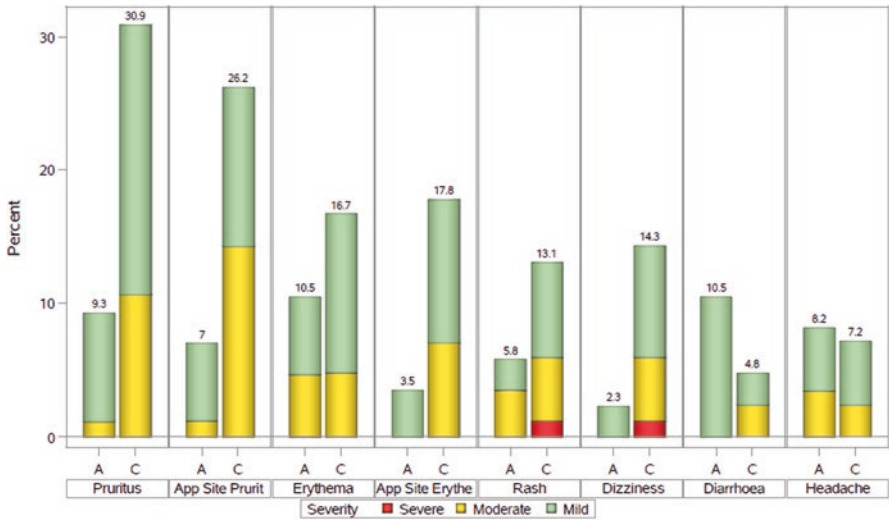
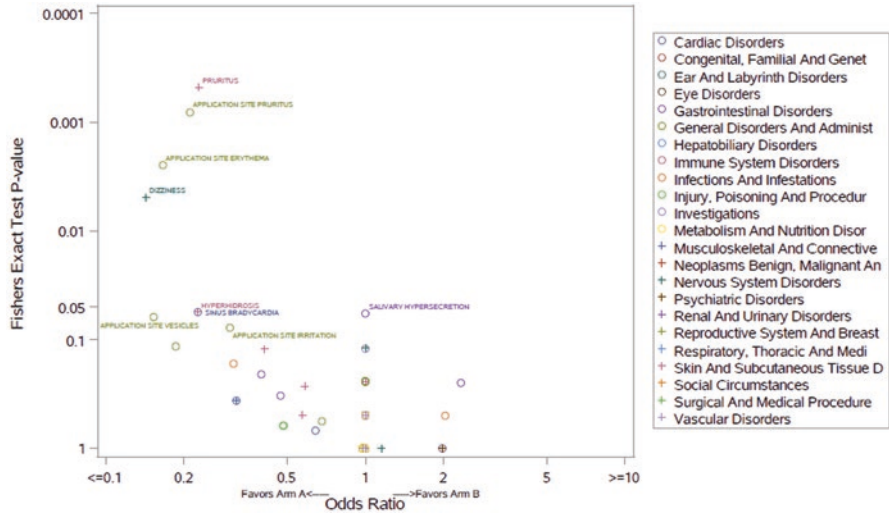


Fig. 2 Adverse events by Preferred Term

A plot of adverse event SOC in descending frequency, by treatment and indicating maximum severity, is a helpful way to quickly show results to the DMC as seen in Fig. 1.

This plot can similarly be done for PTs. This would take many pages, so filtering likely would be done. This could filter to only include the most frequent PTs overall,



**Fig. 3** Adverse events volcano plot

most frequent PTs in a particular treatment, etc. Figure 2 shows this sorted on the most frequent PTs overall.

A volcano plot can be very helpful place for DMC members to start their AE review. As seen in Fig. 3, it shows odds ratio (typically on log-scale) on the x-axis and p-value (typically on log-scale) on the y-axis. PTs with a p-value less than 0.05 are highlighted for additional discussion. It’s important to note that there may be events flagged with p-value less than 0.05 that are not of interest (statistically or clinically), and there may be events that have a p-value greater than 0.05 that are of great interest (a 0 vs. 4 comparison on anaphylaxis, for example). The volcano plot is also of most use when just comparing two treatments – it would be difficult to show three or more distinct treatments on this plot.

A plot showing rates of SOC and relative risk by treatment within each SOC is another figure that DMC members gravitate toward, as seen in Fig. 4. Sorting can be done in different ways. The example below sorts by upper limit of confidence interval of relative risk. This is not to imply any statistical significance if the upper limit of the confidence interval is below 1, but again is acting as a filter to help facilitate further discussion. Versions of this could be done on PTs as well, filtering on most frequent or those with most difference (by HR, or by upper or lower confidence limit of HR).

Note that not all deaths will be AEs. For examples, deaths that are more than 28 days after the last dose might not be entered as an AE. And some studies have defined in the protocol that deaths due to disease progression are not entered as AEs. Deaths should be summarized for the DMC, but this information might be in two locations – one from the AE data, and one from a different data source of all deaths, or from the end of study or disposition data. The summary of death may show categorized reason for death, and may indicate which were within, say, 28 days of last

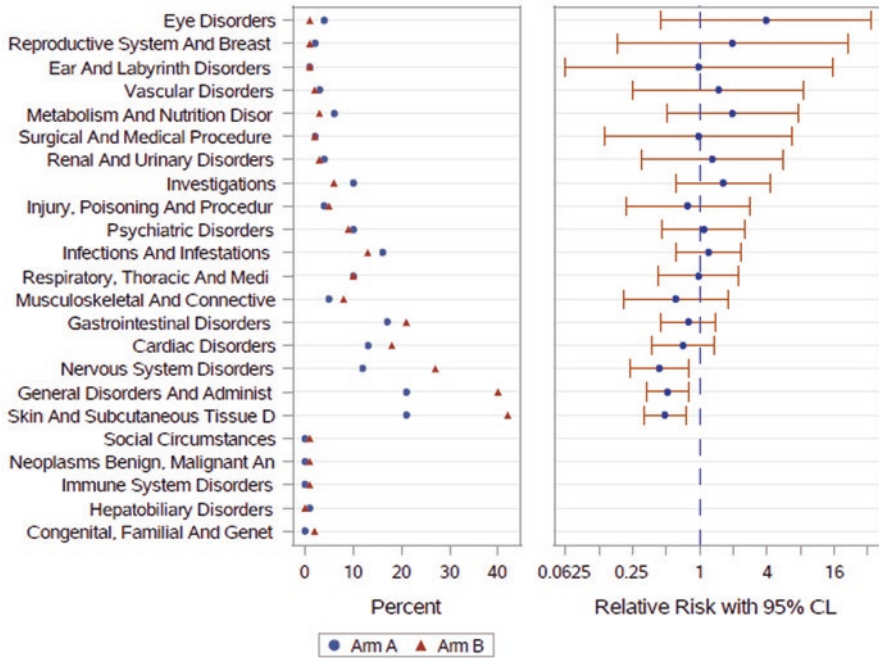


Fig. 4 Adverse events dot plot and relative risk

dose and which were beyond that time frame. A listing of deaths is commonly included as well.

Laboratory data is commonly provided, although this may not be as useful to the DMC as the adverse event data. The DMC may be more focused on laboratory abnormalities that have clinical consequence, in which case those will be captured in the AE data. Laboratory data is commonly categorized as either normal or as abnormal on a Grade 1–5 scale. Some lab parameters have a grading scale in two directions – one for the low (“hypo”) values and one for the high (“hyper”) values. The DMC may be interested in just one or both directions for these lab parameters. It is very easy and common mistake to have long, but unhelpful continuous summaries of lab data – repeating summaries for pages and pages for each time point within every lab parameter. The DMC must be provided more helpful outputs. If needed, the DMC can request additional materials from the SDAC.

A simple approach is to summarize maximum post-baseline grading for lab parameter (including “hypo” and “hyper” summaries separately). This is a short output and distills the most important features – looking to see if one arm or another has an excess of worst grades. A helpful figure for the lab data is a box-and-whisker plot which also includes means over time, as seen in Fig. 5. One figure per lab parameter is quick to review and shows visually both the overall trends (mean over time) as well as extreme values (points outside the whiskers). It is common to show a second figure per lab parameter representing the change from baseline. The

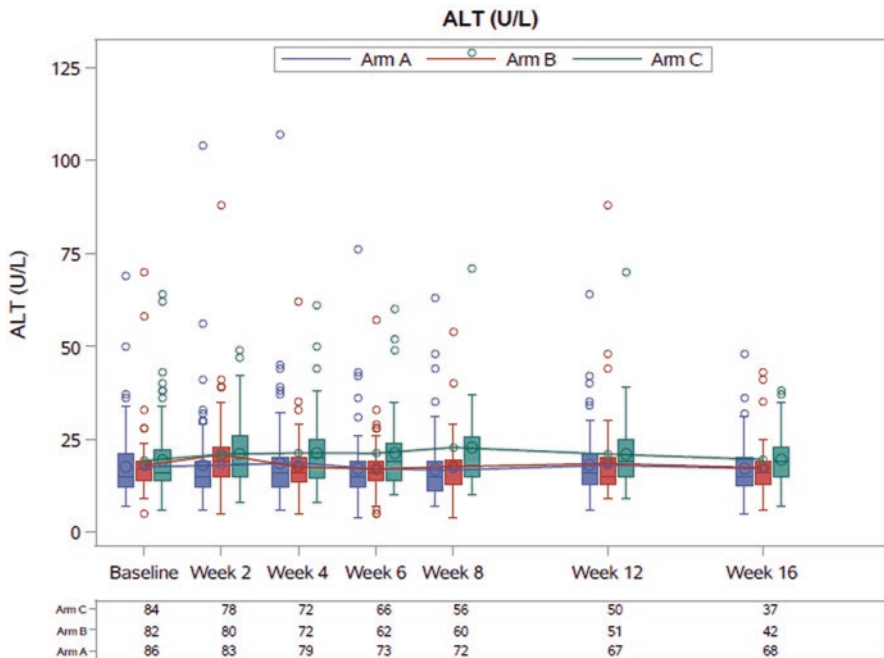


Fig. 5 Laboratory values over time

box-and-whisker plots likely will only include results from nominal visits, not any unscheduled visits. However, the table of maximum post-baseline grading will include all visits, including unscheduled visits.

If lab data is summarized in a table over time as continuous data, include change from baseline results and ensure outputs are in consistent units, using the units expected by the DMC members (which might be SI units, or might be U.S. conventional units). A summary over time might also include categories for values that would cause the DMC to have additional discussion.

Liver function tests (LFTs) – including ALP (alkaline phosphatase), ALT (alanine transaminase), AST (aspartate transaminase), and bilirubin – are a particular concern of DMCs because many treatments are known or suspected to cause hepatotoxicity. It is very common to have a distinct table summarizing number of subjects who have at least one ALT ≥3xULN, ≥5xULN, etc. eDISH (evaluation of drug-induced serious hepatotoxicity) plots are a convenient way to graphically assess ALT and AST vs. bilirubin values. Values are assessed standardized compared to multiples of upper limit of normal (ULN). Both the table and figure will help the DMC to assess if Hy’s Law laboratory criteria have been met (ALT or AST ≥3xULN simultaneously with bilirubin ≥2xULN).

Figure 6 shows the maximum post-baseline AST, ALT, and ALP vs. maximum post-baseline bilirubin. Each subject is only included once. A value in the top-right quadrant for AST and ALT plots might meet Hy’s Law laboratory criteria. However,



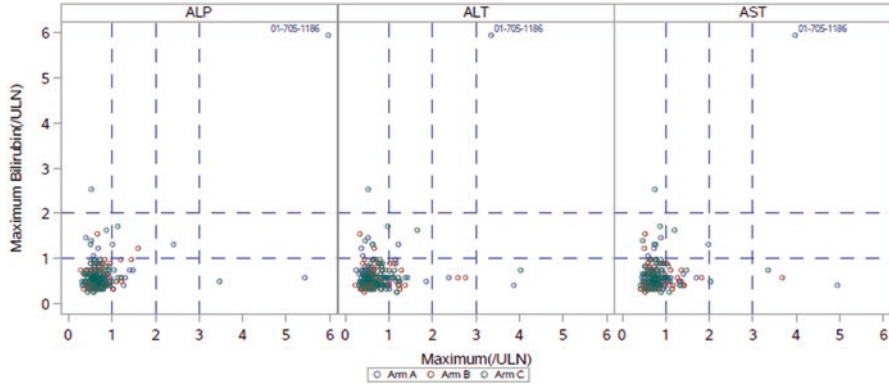


Fig. 6 eDISH plot

there is a chance that values in the top-right quadrant reflect elevations that were not synchronous.

A similar figure could be created that includes every visit. However, matching up ALT and AST visits vs. bilirubin visits to ensure they were synchronous is not always trivial if there are repeat assessments at a visit or unscheduled visits. It seems more common to present the maximum values of the parameters, and then investigate the specific patients of interest – those in the top-right quadrant – to see if elevations were synchronous.

If there are a small number of subjects with LFTs of interest (e.g., have met laboratory criteria for Hy’s Law), a patient profile plot can be helpful, as seen in Fig. 7. These track multiple lab parameters over time (relative to multiples of upper limit of normal for each parameter). It can easily be seen if elevations are synchronous, and if elevations persist or are short-lived.

Looking at shifts from baseline to maximum in a table is helpful (perhaps looking at maximum toxicity grade vs. baseline toxicity grade). But a figure can also be instructive, as seen in Fig. 8. Here’s an example looking at baseline vs. maximum value and highlighting subjects who have more than a 2xULN maximum. It’s easy in this plot to see if these subjects are in the top-right corner of the plot which would indicate being abnormal at baseline, compared to the top-left corner of the plot which would indicate a new lab toxicity.

Listings of lab data can become too long very quickly. A helpful approach is to only list results that are Grade 3 or higher. Include other results from that lab parameter for that subject as well, so that the DMC can easily see the values that preceded and followed the high-grade lab result. Highlight (in bold or in a different color font) the high-grade value that triggered the patient’s lab parameters being included in the listing.

Vital signs are not usually of interest unless the study is specifically intended or known to impact systolic blood pressure (SBP), diastolic blood pressure (DBP) or heart rate (HR). If of interest, include summaries that are similar as for lab data. Summarize the number of subjects who have had at least one value of certain

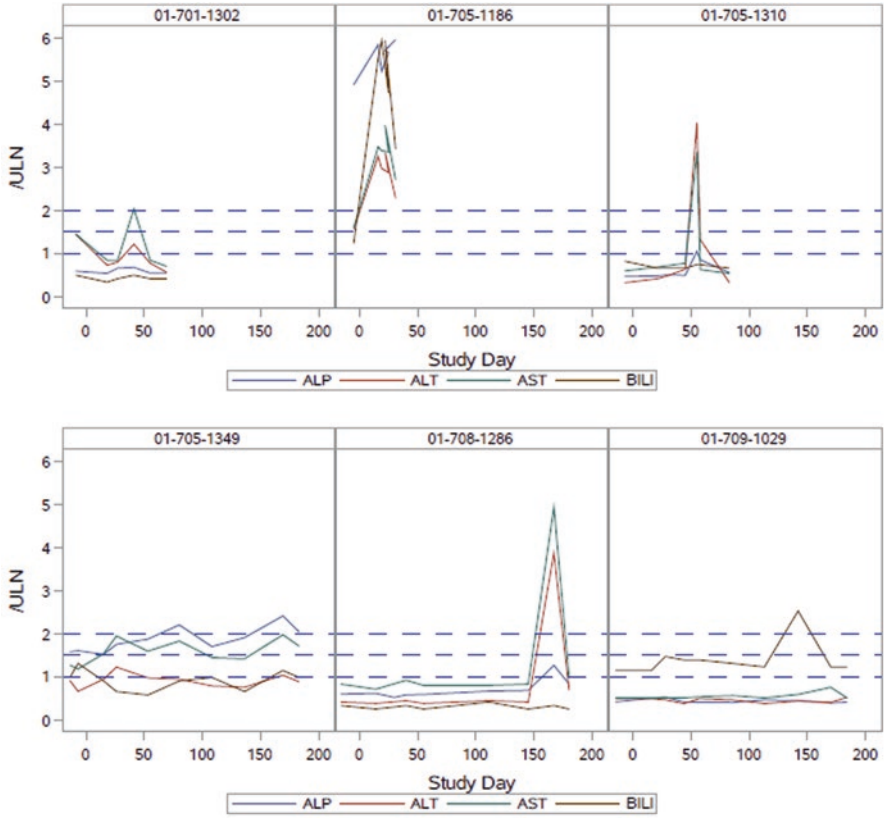


Fig. 7 Laboratory values over time by patient

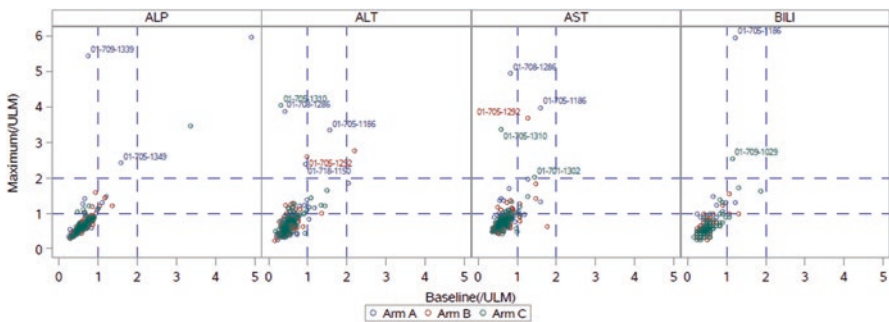


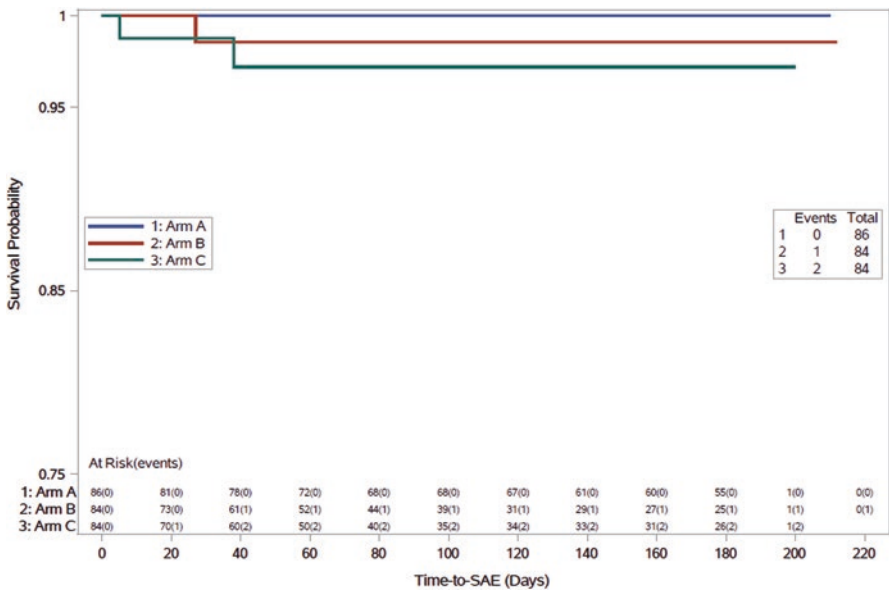
Fig. 8 Laboratory values baseline vs. maximum post-baseline

critical thresholds (e.g., SBP > 180 mmHg with an increase >20 mmHg from baseline) and include a box-and-whisker plot of SBP, DBP, and HR over time. Temperature and weight are typically not an informative way to address any safety

concern, although those outputs may yield interesting results (e.g., increasing weight is a sign of efficacy in studies of patients with Crohn’s Disease, and short-term summary of temperature might be of interest in a vaccine study). But in general, summaries from the AE outputs of terms such as “Weight decreased” or “Pyrexia” or similar would yield more informative safety results than from the vital signs dataset.

Other data might be included as needed for the study (e.g., QTc, ECG). The sponsor and DMC should always remember though that the DMC outputs do not need to include every piece of data collected. Generally, a summary of AEs (and SAEs in particular) will suffice instead of including tertiary safety parameters.

Kaplan–Meier figures are commonly used for presenting efficacy data where the endpoint is time-to-event data where some subjects have experienced the event and others are censored without yet having experienced the event. This is commonly seen for endpoints such as time to death, or time to disease progression (or death). However, Kaplan–Meier figures can also be used to represent safety data in helpful way for the DMC to reveal information about the time pattern of the events. For example, a Kaplan–Meier figure of time to first serious adverse event, as seen in Fig. 9, helps reveal if these SAEs are primarily early in treatment or evenly spread out over time and that could impact DMC recommendations on how to reduce the



**Fig. 9** Time to first Serious Adverse Events

risk of these in the future. The example below has only had three SAEs in the study, but might be more informative as the study matures.