Analysis and Displays Associated with Safety Topics of Interest: Focus on Phase II to IV Clinical Trials

phuse.eu

Contents

List of Figures

List of Tables

1.0. Overview: Purpose of this Document

This white paper is intended to provide recommendations on analyses and displays to sponsors who are planning analyses for safety topics of interest for Phase II to IV clinical trials and integrated summary documents (or other documents in which analyses of safety are of interest). The recommendations pertain to analyses that help clarify the salient features of an adverse event (e.g. whether the adverse event occurs early versus late versus randomly over time) or help reviewers determine whether an event should be considered an adverse drug reaction (ADR; defined as reasonably likely to be caused by study drug).

We consider "safety topics of interest" a catch-all term that includes adverse events of special interest (AESIs), identified or potential risks that need to be further characterised, potential toxicities that all products should consider, potential findings based on drug class, or topics anticipated to be requested by a regulatory agency for any reason. Depending on the context, we may refer to a specific safety topic of interest as "adverse event of special interest", "adverse event of interest", "adverse event", "event of interest" or simply "event" or "safety topic". In some cases, the topic of interest could already be considered an ADR.

During the life cycle of a drug, safety topics of special interest may emerge from several sources. Examples include:

- Toxicology and other nonclinical data that may suggest potential toxicities in humans
- Known class effects
- Adverse effects described in literature
- Post-marketing data that suggest an AE could be an ADR, or an ADR that appears more frequent or severe than is listed in the product label
- Toxicities observed in Phase I to IV clinical trials. These may be generated by observing a single event, or by observing an imbalance in an aggregate analysis that disfavours the drug.
- Regulatory requests
- Review of data for safety reports, e.g. PSUR, DSUR.

This white paper is a follow-on to the PHUSE AE white paper that provides recommended tables, figures and listings (TFLs) from clinical trial data for adverse events with a focus on general safety signal detection [1].

This white paper focuses on additional (or, in some cases, different) TFLs that will likely be helpful for ADR determination, ADR characterisation or ADR communication purposes. Not all safety topics of interest will require additional summaries or analyses. In some cases, the summaries and analyses provided as part of safety signal detection will be sufficient. However, additional TFLs are often required to address topics such as time to onset, underlying risk factors, persistence/transience, reversibility and dose dependency, consistent with the suggestions provided in the FDA Safety Reviewer Guidance [2, pp. 24–25, 44–49] and FDA Clinical Review Template [3, Section 7.5]. In some cases, different TFLs are required for special topics of interest where more general approaches do not apply (e.g. when number of events is more useful than number of patients with an event). Education and communication around useful TFLs will lead to improved and harmonised product life cycle management across therapeutic areas by ensuring that reviewers receive clinically relevant and meaningful analyses. This white paper reflects recommendations that would lead to more consistent TFLs, but the recommendations should not be interpreted as "required" by any regulatory agency.

Another purpose of this white paper (along with the other white papers from the project team) is to improve expertise in safety analytics across the multiple disciplines involved with planning, interpreting and reporting safety analyses. Statisticians can and should assist cross-disciplinary teams with creating analytical plans that are consistent with sound statistical principles. They should also assist with the interpretation of results. This assistance is important, even when inferential statistics are not used. Identifying an appropriate method is especially a concern when multiple studies are combined (e.g. via poor pooling practices for integrated summaries) (Section 7.1.3 of Attachment B of the FDA Clinical Review Template [3]). Safety physicians have often relied on qualitative analyses of case reports, looking at individual or small clusters of events. Recently, there has been an increased emphasis on aggregate reviews of safety data. As noted in Section VI of the CIOMS Working Group VI report [4], while medical judgement remains critical in the interpretation of safety data, descriptive and inferential statistical methods can help medical personnel decide whether chance variation is a possible explanation for what is observed or whether it is more likely that some genuine drug effect has occurred. This requires statisticians to increase their engagement and help cross-disciplinary safety management teams to think more quantitatively. This also requires nonstatistical disciplines to obtain a higher level of analytical knowledge.

This document is focused on a high-level overview of tables and figures that can be useful for understanding safety topics. Having strong statistical guidance throughout the planning and review of safety data is very important as many of the statistical methods that are presented in tables require solid understanding of the underlying methods and assumptions, as well as the data limitations, in order to implement and use them appropriately.

2. Problem Statement

Industry standards for data collection and storage have evolved over time: Clinical Data Acquisition Standards Harmonization (CDASH), observed data (SDTM), and analysis datasets (ADaM). However, there is a lack of a commonly agreed systematic approach to identify and analyse safety signals from clinical trials [5]. White papers outlining recommendations to identify safety signals have been created [1, 6, 7], but none cover recommendations for additional summaries that might be useful when a safety signal has been identified and requires further characterisation. Additionally, the analyses recommended for general safety assessment may not apply to safety topics of interest. This white paper is intended to fill that gap by providing example displays that are often useful for safety topics of interest.

3. Scope

This white paper pertains to adverse event data that are collected via passive or active solicitation. In most cases the adverse event is translated into a medical dictionary. Currently, the most common medical dictionary for clinical trials is MedDRA. Special topics in which data come from specialised collection methods (e.g. C-SSRS for suicidal thoughts and behaviours) are out of scope. Specific recommendations for laboratory measurements (in general), hepatotoxicity, vital signs and electrocardiogram measurements will be addressed in separate white papers, so are out of scope for this white paper.

The focus of this white paper is primarily on Phase II to IV clinical trials, though some of the content may apply to Phase I studies or other types of medical research such as observational studies.

Detailed variable specifications for TFLs or dataset development are out of scope. Important concepts such as "safety population" (the set of patients in a clinical trial that will be included in safety analyses in general), and "risk set" (the subset of the safety population included in a particular analysis) are out of scope. We encourage code developers to use the concepts outlined in this white paper as the basis for dataset, TFL and/or interactive tool specifications.

4. Acronyms and Definitions

Acronyms

Term	Description
ADaM	analysis data model
ADR	adverse drug reaction
AE	adverse event
AESI	adverse event of special interest
CDASH	Clinical Data Acquisition Standards Harmonization
CI	confidence interval
CIOMS	Council for International Organizations of Medical Sciences
CSR	clinical study report
C-SSRS	Columbia-Suicide Severity Rating Scale
CTCAE	common terminology criteria for adverse events
EAIR	exposure-adjusted incidence rate
EAER	exposure-adjusted event rate
FDA	Food and Drug Administration
FMQ	FDA MedDRA query
HLGT	high level group term
HLT	high level term
ICH	International Conference on Harmonisation
IRR	incidence rate ratio
LLT	lowest level term
MACE	major adverse cardiovascular events
MedDRA	Medical Dictionary for Regulatory Activities
PBO, PL	placebo
PSAP	program safety analysis plan
PT	preferred term
SAP	statistical analysis plan
SDTM	study data tabulation model
SOC	system organ class
SMQ	standardized MedDRA query
TEAE	treatment-emergent adverse event
TFLs	tables, figures, and listings
TRT	treatment
WHO	World Health Organization

Definitions

The following provides definitions for terms used in this white paper. As discussed in the PHUSE white paper on adverse events [1], nomenclature varies among sources and there is no agreement. There are examples where multiple terms are used for the same quantity and other cases where the same term is used for different quantities. Thus, it is important to include clear definitions in statistical analysis plans (SAPs), program safety analysis plans (PSAPs), clinical study reports (CSRs), submission documents and TFLs (e.g. in footnotes if needed). When reporting percentages, incidence or rates, it should be clear whether the numerator is the number of patients or the number of events and whether the denominator is the number of patients in a population, the time at risk or the total time exposed. As noted in PHUSE 2017 [1], we concur with sources that include "person-time" or "exposure-adjusted" in the name for clarity when time is used in the denominator.

Adverse event (AE): Any untoward medical occurrence associated with the use of a study drug in humans, whether or not it is considered study-drug-related. See Section 1.2. of Attachment B of the FDA Clinical Review Template [3]. See also Section 10.2 in PHUSE 2017 [1].

Adverse drug reaction (ADR): An undesirable effect, reasonably likely to be caused by a study drug, which may occur as part of the pharmacological action of the study drug or may be unpredictable in its occurrence. See Section 1.2. of Attachment B of the FDA Clinical Review Template [3] and the FDA Guidance on Safety Reporting Requirements for INDs and BA/BE Studies [8]. See also Section 10.2 in PHUSE 2017 [1].

Percent (in the context of AE reporting): The number of patients with an event divided by the number of patients at risk for the event, multiplied by 100. As discussed in the PHUSE AE white paper [1], this is sometimes referred to as the event rate, incidence rate, crude incidence rate, incidence proportion or cumulative incidence.

Exposure-adjusted event rate (EAER): The number of events (if a patient has more than one occurrence of the same event, all occurrences are counted) divided by the total time exposed [9]. This is sometimes referred to as person-time absolute rate [10]. Total time exposed is calculated as the sum of each patient's time in the interval, whether or not the patient experienced the event. The time unit used can be changed (e.g. if the original units are events per person-year, this can easily be converted to events per 100 person-years by multiplying by 100). The exposure time should be based on the same time interval in which any events that occur would be counted.

Exposure-adjusted incidence rate (EAIR): The number of patients with an event divided by the total time at risk for the event [9]. Total time at risk will be calculated as the sum of time from the first dose (or randomisation) to first event for

patients who experienced the event and the time during the entire assessment interval for patients who do not experience the event. This is sometimes referred to as the incidence rate or person-time incidence rate [10, 11]. As noted above, we believe the addition of "exposure-adjusted" or "person-time" is beneficial for clarity.

Safety topics of interest: A catch-all term that includes adverse events of special interest (AESIs), identified or potential risks that need to be further characterised, potential toxicities that all products should consider (e.g. hepatic-related events), potential findings based on drug class, or topics anticipated to be requested by a regulatory agency for any reason. Depending on the context, we may refer to a specific safety topic of interest as "adverse event of special interest", "adverse event of interest", "adverse event", "event of interest" or simply "event" or "safety topic".

Static display: A TFL that is created without any interactive features, to be viewed in its entirety on one or more pages. This is in contrast to an interactive display that allows point-and-click technology and/or scroll bars to see additional data online.

Study-size adjusted percentage: A percentage that is calculated when multiple controlled studies are combined. The percentage in each treatment arm is calculated by weighting the observed percentage within a study by the percentage of patients in that study among the pooled population [12, 13]. This is also referred to as a study-size-adjusted incidence percentage. See Section 10.1 in PHUSE 2017 [1] for an example (Table 10.1).

Treatment-emergent adverse event (TEAE): An AE following the first administration of the intervention that is either new or a worsening of an existing AE. While this definition appears straightforward, there are multiple ways in which this is implemented across the industry. See Section 10.2 in PHUSE 2017 [1].

5. Background

The PHUSE Computational Science Collaboration is an initiative involving PHUSE, the FDA and industry that identifies computational science priorities that could be addressed by collaboration, crowd sourcing and innovation [14]. Several Working Groups have been created to address many of these challenges. The Standard Analyses and Code Sharing Working Group has led the development of this white paper, along with the development of a platform for creating and storing shared code.

Several existing guidance documents [2–4, 8, 15–25] contain suggested TFLs and/or discussions around analyses for common assessments and quantitative measurements. These documents were used as a starting point in the development of this white paper. Additional references used to inform recommendations have been cited throughout the document.

Members of the Analysis and Display White Papers Project Team reviewed these documents and shared ideas and lessons learned from their experience. A draft of this white paper was developed and posted in the PHUSE environment for public comments.

Most contributors and reviewers of this white paper are industry statisticians, with input from non-industry statisticians (e.g. the FDA and academia) and industry and non-industry clinicians. Additional input (e.g. from other regulatory agencies, the ICH, the World Health Organization [WHO]) for a future version of this white paper would be beneficial.

6. General Recommendations

This section contains general recommendations for analyses and displays that apply to safety data.

6.1 Choice of Comparative Metric for Incidence Proportions/ Percentages

Establishing the risk profile of a drug requires a comparison of AE data of patients taking the investigational product to a control (either placebo or active control). As noted by Zhou et al. [9]:

> The metrics for measuring treatment difference could be on an absolute scale (e.g., risk difference for binary endpoints, difference in median survival time for time-to-event variables) or a relative scale (e.g., relative risk, odds ratio and hazard ratio). An appropriate metric is important to accurately quantify and communicate the increased risk and thus inform the treatment decision for patients. However, no clear guidance on which metric to use in the safety assessment seems available.

There are advantages and disadvantages across these choices [5, 9, 26]. As noted by Zhou and colleagues [9], the risk difference can more directly reflect the magnitude of patients that could be affected by a risk. It's also an easier metric to implement for events with very low frequency. However, relative metrics are sometimes used as a flagging mechanism for identifying events that may warrant further investigation (e.g. identifying events occurring at least two times as frequent with drug versus placebo).

When a relative metric is of interest, it's tempting to choose the relative risk as it's easier to understand than the odds ratio. However, there are advantages to the odds ratio that are not well understood. The risk ratio has problems as the underlying percentage gets larger. Think, for example, of an event with a percentage in the control arm of 50%. In that case, the risk ratio can never be higher than 2. However, the odds ratio can still be small or large, which makes it more effective for ascertaining the magnitude of difference between treatment groups. Odds ratios have better mathematical properties than risk ratios. This is partly because the log odds ratio can take values from – infinity to + infinity, regardless of the value in the control denominator (assuming the control denominator is non-zero – both risk ratio and odds ratio are problematic if there are zero cells). It also is invariant to coding changes, i.e. if you count the percentage of patients with an event vs. those without, the odds ratio will just be the reciprocal. This is not true for relative risks. You will get different relative risks (not just the inverse) and different inferential statistics depending on whether you count patients with a particular event or without it.

In summary, the risk difference is excellent for understanding

the public health impact of an event, but not as good for understanding the relative impact, and its use makes it more difficult to identify rare events that might require further investigation. Risk ratios are easily interpretable and good for understanding the relative impact, but not as good for understanding the public health impact, and their use makes it more difficult to identify events that require further investigation when the background rate is large. While odds ratios are not as easily interpretable, they are good for understanding relative impact and can more easily be used as a flagging mechanism regardless of background rate. Therefore, odds ratios can be useful for signal detection. For presentation to the public, there are distinct advantages to presenting absolute differences or risk ratios rather than odds ratios. The recommended displays given in this white paper show particular comparative metrics (generally based on absolute difference); however, these are not intended as a recommendation. Companies may choose different metrics if desired. When interactive displays are created, we encourage developers to include the ability for the user to choose among multiple metrics. For AESIs, providing multiple metrics (e.g. one absolute metric and one relative metric) is often warranted.

6.2. P-values and Confidence Intervals

There has been ongoing debate on the value (or lack of value) of the inclusion of p-values and/or confidence intervals in safety assessments [27]. This white paper does not attempt to resolve this debate. As noted in the FDA Clinical Review Template (e.g. Section 7.4.2 of Attachment B) [3], p-values or confidence intervals can provide some evidence of the strength of the findings, but unless the trials are designed for and powered appropriately for hypothesis testing, these should be thought of as descriptive. The International Conference on Harmonisation E9 Section 6.4 [28] notes that descriptive statistics are generally used for safety, with confidence intervals wherever it aids in interpretation. The International Conference on Harmonisation E9 [28] also mentions that p-values are sometimes useful as a flagging mechanism to highlight differences worth further attention. Ma and colleagues [5] state that it's important to show some measure of uncertainty (confidence intervals, p-values, posterior credible intervals or posterior probabilities). Throughout this white paper, confidence intervals (not p-values) are included in several places as they do provide at least a crude estimate of the strength of evidence. Where these are included, they should not be considered as describing a hypothesis test. If a sponsor or compound team decides p-values would be helpful, they can be added. If p-values are added, we recommend the actual p-values be reported instead of an asterisk indicating when a threshold is met. This is more consistent with the idea of using it as a tool for interpretation (by knowing relative strength of evidence among events) instead of a hypothesis test, as emphasised in a February 2016 statement from the American Statistical Association [29]. (Note that subsequently The American Statistician journal published an entire special issue (March, 2019, *https://tandfonline.com/toc/utas20/73/sup1*) dedicated to the topic of p-values. See Wasserstein et al. [30] for high-level advice and a summary of the articles that are published in the special issue.

Some teams may find p-values and/or confidence intervals useful to facilitate focus but have concerns that high p-values or confidence intervals including 0 (for differences) or 1 (for ratios) may lead to unwarranted dismissal of a potential signal. Conversely, there are concerns there could be misinterpretation of p-values adding potential concern for too many outcomes. Similarly, there are concerns that the lower or upper bound of confidence intervals may be misinterpreted. As noted in ICH E9 Section 6.4 [28], the considerable imprecision that arises from low frequencies of occurrence is clearly demonstrated when confidence intervals are used. For example, a wide confidence interval would have a very high upper bound. Sometimes, seeing a very high upper bound causes undue alarm that a risk could actually be that high, when in fact there are simply too few cases to make an estimate. It is important for those interpreting the TFLs to be educated on these issues. When reporting safety information, it's useful to include within-arm descriptive statistics, as well as a measure of the difference between arms (see Section 6.1), along with a confidence interval for the measure of difference.

When p-values or confidence intervals are used, the decision to conclude that any given AE is an ADR should never be based solely on a p-value or confidence interval [31]. The decision should be based on the totality of evidence coming from various sources of evidence, based on levels of evidence and use of medical judgement. In general, two practical frameworks to identify ADRs are used: 1) the CIOMS Working Group [32] and 2) Bradford Hill criteria [33, 34]. These frameworks are flexible and help think through the various pieces of information such as frequency of adverse event, timing of occurrence, pre-clinical findings and mechanism of action. When only percentages are included in displays (e.g. no p-values, no confidence intervals, no risk difference, no risk ratio, no odds ratio), it may be difficult to determine which events warrant further scrutiny. Unless the sponsor or compound team determines some other objective process, the review of AEs would completely rely on the individual physicians and their way of processing the information.

6.3. Importance of Visual Displays

Communicating information effectively and efficiently is crucial to enable rational decision-making. Current practice, which focuses on tables, has not always allowed us to communicate information effectively since tables and listings may be very long and repetitive. Graphics, on the other hand, can provide more effective presentation of complex data, increasing the likelihood of improving the ability to make clinical decisions [35, 36]. They can also facilitate identification of unexpected values.

Standardised presentation of visual information is encouraged. While this white paper focuses on static displays, we do include some notes for areas where interactive visual capabilities would be beneficial. The displays in this white paper can serve as a source for interactive safety review packages. It provides the types of information that would be of interest to include in a package, and analytical considerations that would be important no matter how the information is displayed.

6.4. Integrated Analyses

As noted in Section 10.1 of PHUSE (2017) [1], for submission documents, TFLs are generally created using data from multiple clinical trials. Determining which clinical trials and which treatment arms to combine for a particular set of TFLs can be complex. Section 7.4.1 of the FDA Reviewer Guidance [2] contains a discussion of points to consider. For purposes of this white paper, we assume not all studies will have the same

doses and that all doses of the investigational study drug that fall within the range of draft label dosing will be included as a single treatment arm. However, the TFLs can be adapted to different scenarios. Generally, when calculating summary metrics (e.g. odds ratio, risk ratio, risk difference), confidence intervals, and/or p-values, incorporating a method that accounts for the inclusion of data from multiple studies (e.g. including study as a stratification variable) is important. When the treatment-placebo randomisation ratio (after pooling of any dose groups) is not constant across the studies included in the integrated summary and only crude percentages are calculated, then the review of data is subject to potential misinterpretations (e.g. Simpson's paradox [13]). Creating visual displays or tables in which comparisons are confounded within study is discouraged.

Assessing results within each study, in addition to assessing the pooled results, is always a good idea.

Furthermore, we emphasise that it is always important to discuss the SAP for the integrated data with regulatory review divisions, in order to ensure that the sponsor and regulatory review division are aligned on the utility of the planned analyses. (See Sutter, 2019 [37] for more information.)

6.5. Competing Risks

Competing risks are events that preclude the occurrence of the main event of interest. For example, if the event of interest is myocardial infarction, death from other causes would be considered a competing risk. Competing risks are different from other concurrent events in that they actually preclude the event of interest from happening, whereas events like early study discontinuation prevent the event from being observed. Various authors (e.g. [38–43]) have written about the need to consider competing risks in the assessment of the risk/probability of adverse events.

A full treatise on this topic is not possible; nevertheless, we offer two main insights. First, if interest is in determining whether or not the drug is causally related to the AE, standard Kaplan-Meier or Cox proportional hazards methods perform better than methods that take into account the competing risk [44]. Alternatively, if interest is in getting an accurate percentage of patients with the event (e.g. for the label), estimation methods that take competing risks into account will be useful.

7. Tables and Figures for Individual Studies

7.1. Recommended Displays

Throughout this section, recommended displays are provided. Where displays are applicable to controlled data, examples include two treatment arms (low dose and high dose) and a placebo arm. The tables can be modified as needed under different scenarios. A separate white paper titled "General Output Tips and Considerations" is planned, which will provide recommendations for headers and footers, among other topics.

7.1.1. Patient profile

It is often important to examine a range of data types for each participant in order to understand AEs and their interrelationships with other AEs and other types of data. This may be needed for a subset of patients with a particular event but at times may be required for all patients. Textual patient profile listings and/or graphical displays may be used. Graphical displays are particularly useful for clinical review when continuous endpoints such as labs or vital signs are included. Graphical patient profiles with treatment course, AEs (with toxicity grade), concomitant medications, relevant laboratory findings and medical history (e.g. pre-existing conditions) greatly help with the interpretation of the treatment relatedness of AEs. For safety topics of interest, we recommend the patient profile to be tailored to the event of interest, e.g. based on available data, medical judgement and review of literature on risk factors or potential confounders. There is a lot of information collected for each participant, and not all of it is relevant for the particular safety topic of interest. If there is additional information collected for the topic, the relevant information should be included in the patient profile. For example, if the topic is suicide and Columbia-Suicide Severity Rating Scale (C-SSRS) measures were collected, these data should be included in the patient profile. Similarly, the patient profile does not need to include data that are not important for the topic of interest.

Figure 7.1 shows a patient profile with two lab parameters displayed in separate panels. Some situations may permit multiple lab parameters to be displayed together in a single panel either in their shared International System of Units/conventional unit or as a fraction of the respective ULN. Values that exceed the ULN are annotated for each parameter in this display as an example of how clinically important values can be highlighted for the reviewer. The lower part of the graph shows the participant's drug exposure, adverse events and concomitant medications as segmented lines. The lines are ordered by start date of the first occasion. For example, the line for "AE #1" is before the line for the second concomitant medication, "Con med #2", as the first occurrence of "AE #1" started before "Con med #2". The colour shows the adverse event severity level as the CTCAE grade. (Another possibility for showing severity is the "Mild", "Moderate" and "Severe" scale.) The website *www.colorbrewer2.org* is great for helping to choose appropriate colours. Relevant demographics, baseline characteristics, dosing dates, etc. are included, as in the lower left panel of the example. By presenting the time course of changes in continuous parameters in context with study drug dosing, AEs and other medications, a graphical patient profile can depict the full participant experience and facilitate clinical review.

We have shown a simple example of a graphical patient profile. In many cases, it will need to be more complex in order to capture the information needed to understand the topic of interest.

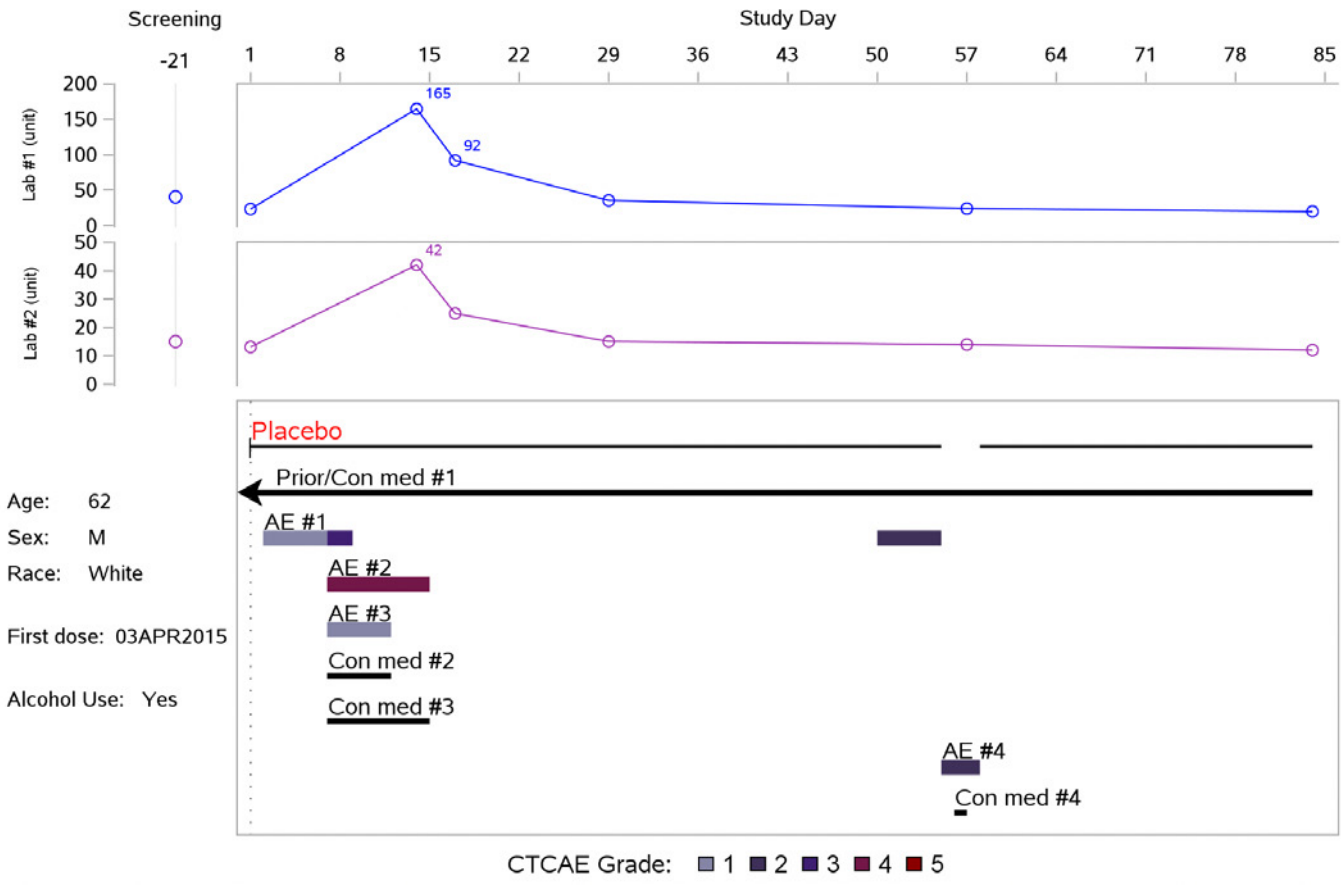
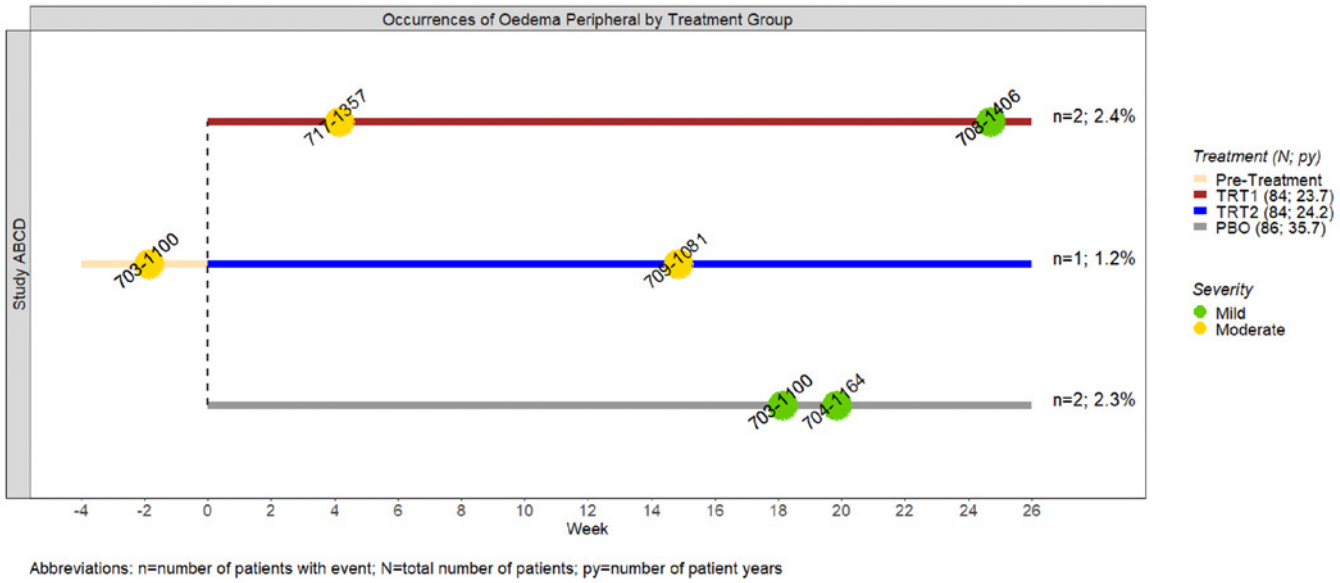


Figure 7.1. Graphical Patient Profile

7.1.2. Events with Very Low Frequency

Many events are so infrequent that the focus of assessment will be on case-level reviews rather than summary tables. In this case, creating good patient profiles (possibly tailored to the event of interest, as noted in Section 7.1.1) is recommended. Creating listings instead of graphical patient profiles can provide the information for case reviews, but is generally more difficult to see relationships of the event with various factors (e.g. concomitant medications) over time. Additionally, a figure that displays the study design with symbols showing where each event occurs can be considered (see Figure 7.2). This might be especially helpful for studies with complex study designs. It's an easy way to see where events are occurring across treatment dose arms and the different periods across the study.



Abbreviations: n=number of patients with event; N=total number of patients; py=number of patient years

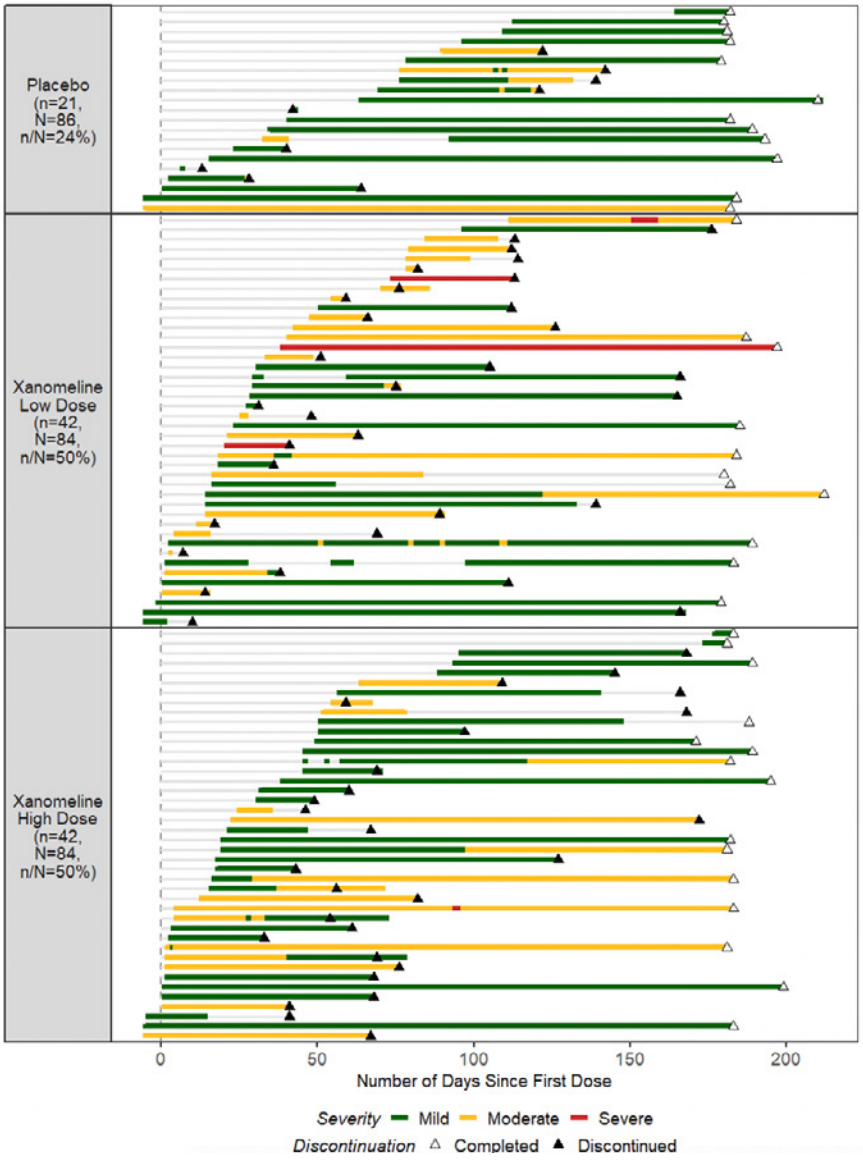
Figure 7.2. Patients with Event XXXX. Figure that displays the study design with symbols showing where each event occurs and the associated patient ID.

7.1.3. Time Course of an Event

7.1.3.1. Onset, Duration

For most safety topics of interest, the time course of an event is of interest, either to help determine whether the event is an ADR and/or to further characterise ADRs. The factors that are generally of interest include event duration, time of onset, frequency of recurrence and changes in intensity over time. The onset/duration plot (Figure 7.3) provides a visual display of an event's (including recurrent events) onset time and duration for each participant along a common time axis. Study drug is depicted by a thin line behind the broader lines for the events, which are coloured according to event intensity. Other relevant times, such as time of study discontinuation or completion, may be represented using symbols. This figure has the advantage of visually displaying individual participant data in a relatively concise manner. Some medical representatives find displays of individual participant-level data useful as it often helps achieve a greater understanding of the data over just having displays with group-level summaries. However, when the number of patients with the event is large, any static version of the onset/duration plot could be difficult to read. It might make sense to include a link from a CSR or integrated summary document to an interactive version of the display with different options for sorting and scroll bars to enable the ability to see the large number of patients with events. A PHUSE Data Visualization project team conducted a pilot in which an interactive display was provided in a submission. For more information on how this was achieved, see [45, 46].

Of note, the specific features that can be included in an onset/duration plot will depend on collection. For example, using colour to indicate event intensity over time can only be considered if changes in event intensity are collected over time (both worsening and improving). Some adverse event collection methods only capture worsening.



Abbreviations: n = number of patients with at least one adverse event; N = number of patients in treatment arm; % = percent of patients in each treatment arm with at least one adverse event. Horizontal grey lines represent time on drug.

Figure 7.3. Onset/Duration Plot, or Event Chart

7.1.3.2. When First Event/Time to First Event is of Interest

Often, the onset date of an AE relative to start of treatment is of interest. If we want to summarise how the risk for the first onset of an event is changing over time, it is convenient and appropriate to treat AEs as time-to-event data [36]. Two graphical displays based on non-parametric survival methods may be particularly useful. First, Kaplan-Meier plots by treatment group of the cumulative incidence of the event of interest are useful (see Figure 10 of Amit et al. [36]). A Kaplan-Meier plot is not only good for helping to determine whether the time of first event onset differs among treatment groups, but it can also help us understand whether the first event tends to occur early or late (by inspection of the slope of the curve). A related curve, the hazard curve [36], is even easier to use for understanding how the rate of first occurrence changes over time. (Another related way is to look at EAIRs in time blocks, as noted in Section 7.1.3.4.) Hazard function plots are very similar to incidence rate plots over time blocks. See Section 7.1.3.4 for an example. Note that Kaplan-Meier plots based on integrated data can be subject to confounding/Simpson's paradox unless special methods are utilised [47].

If occurrence of the first event is of interest, summary statistics for EAIRs and incidence rate ratios (IRRs) are also useful. A mock table of EAIRs is shown in Table 7.1. EAIRs do not allow for evaluating risk as a direct function of time like Kaplan-Meier plots do; however, they can still be quite useful, particularly if there is a differential follow-up time amongst treatment groups. For example, in some neuroscience trials, the placebo-treated groups have much larger dropout percentages than the treated groups (e.g. [48]). An analysis of percentages of patients with a particular event will yield biased estimates. However, an analysis that controls for time on study (or time at risk) has a better chance of yielding an unbiased estimate of the treatment effect for the event of interest. As another example, event-driven studies (e.g. an oncology study evaluating progression-free survival) may result in differing follow-up time for safety endpoints among the treatment groups.

7.1.3.3. When Recurrent Events are of Interest

While standard safety analyses focus on the number of patients with at least one event, for some special topics there is also interest in understanding the recurrent events that patients experience throughout the duration of the trial. For example, many drugs to treat diabetes have hypoglycemia as a special topic of interest. With very long-term follow-up, nearly 100% of patients will have at least one episode of hypoglycemia, which makes the percentage of patients with an event a poor way of discriminating between treatments. If a drug reduced the frequency of hypoglycemia substantially, we would need to look at the total number of events experienced rather than just the number of patients with at least one event, in order to understand this aspect of the safety profile. As noted in Section 10.9 of PHUSE 2017 [1], in this and some other circumstances, statistical methods for handling recurrent events [49] might be needed. One metric to assess this is the exposure-adjusted event rate (EAER) [9]. See Table 7.2 for an example display. As with the exposure-adjusted incidence rate, the EAER assumes the risk for the AE is constant over time. With few events, it can be difficult to check this assumption; however, if sufficient data are available to assess the risk over time, assessing the data in blocks of time may be helpful. See Section 7.1.3.4. Other methods may be needed, but for simplicity we focus on the metric EAER.

As an alternative to assessing EAER (or as an addition), a display that enumerates the number of patients with different levels of occurrence is sometimes useful, depending on the event of interest. See Table 7.3.

A useful visual is the onset/duration display (Figure 7.3) that is used for reviewing participant-level data. All events are displayed in that plot. Another alternate is the mean cumulative function plot [50].

One challenge when summarising recurrent events pertains to event collection. There is substantial variability in collection methods for recurrent events, and the variability may exist within a study across sites. For example, some sites may record every event while others record recurrent events as one event with recurrent in the description (e.g. recurrent headaches). Such variation leads to complexities in implementation and interpretation.

Table 7.1. [Special Topic of Interest] Using a Pre-defined List of MedDRA Preferred Terms – Number of Events and Exposure-adjusted Incidence Rates

Preferred Term	PL (N=XX; PY=xxx) n PYE (EAIR)	T1 (N=XX; PY=xxx) n PYE (EAIR)	T2 (N=XX; PY=xxx) n PYE (EAIR)	T1 & T2 (N=XX; PY=xxx) n PYE (EAIR)	EAIR Difference[a] T1-PL (95% CI)	EAIR Difference[a] T2-PL (95% CI)	EAIR Difference[a] T1&T2-PL (95% CI)
Number of events (all pre-defined preferred terms)	xx xxx(xx.x)	xx xxx(xx.x)	xx xxx(xx.x)	xx xxx(xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #1]	xx xxx(xx.x)	xx xxx(xx.x)	xx xxx(xx.x)	xx xxx(xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #2]	xx xxx(xx.x)	xx xxx(xx.x)	xx xxx(xx.x)	xx xxx(xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #3]	xx xxx(xx.x)	xx xxx(xx.x)	xx xxx(xx.x)	xx xxx(xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #n]	xx xxx(xx.x)	xx xxx(xx.x)	xx xxx(xx.x)	xx xxx(xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)

Abbreviations: EAIR = exposure-adjusted incidence rate; CI = confidence interval; N = number of patients; n= number of patients with at least one event; PY = total patient years; PYE = number of person years at risk for a first event.
Notes: EAIR is 100 times the number of patients with an event divided by total time at risk in years.
Patients may be counted in more than one row.
[a]EAIR difference confidence interval from [insert method].

Table 7.2. [Special Topic of Interest] Using a Pre-defined List of MedDRA Preferred Terms – Number of Events and Exposure-adjusted Event Rates

Preferred Term	PL (N=XX; PY=xxx) n n1 (EAER)	T1 (N=XX; PY=xxx) n n1 (EAER)	T2 (N=XX; PY=xxx) n n1 (EAER)	T1 & T2 (N=XX; PY=xxx) n n1 (EAER)	EAER Difference[a] T1-PL (95% CI)	EAER Difference[a] T2-PL (95% CI)	EAER Difference[a] T1&T2-PL (95% CI)
Number of events	xx xx(xx.x)	xx xx(xx.x)	xx xx(xx.x)	xx xx(xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #1]	xx xx(xx.x)	xx xx(xx.x)	xx xx(xx.x)	xx xx(xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #2]	xx xx(xx.x)	xx xx(xx.x)	xx xx(xx.x)	xx xx(xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #3]	xx xx(xx.x)	xx xx(xx.x)	xx xx(xx.x)	xx xx(xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #n]	xx xx(xx.x)	xx xx(xx.x)	xx xx(xx.x)	xx xx(xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
.....							

Abbreviations: EAER = exposure-adjusted event rate; CI = confidence interval; N = number of patients; n= number of patients with at least one event; n1 = number of events; PY = total patient years.
Notes: EAER is 100 times the number of events divided by total time at risk in years.
Patients may be counted in more than one row.
[a]EAER difference confidence interval from [insert method].

Table 7.3. [Special Topic of Interest] Using a Pre-defined List of MedDRA Preferred Terms – Number of Events in Categories

High Level Term / Preferred term		PL (N=xxx) n (%)	T1 (N=xxx) n (%)	T2 (N=xxx) n (%)	T1 & T2 (N=xxx) n (%)
Patients with ≥1 TEAE Injection Site Reactions	Total patients	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
	Patients with 1 event	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xxx(xx.x%)
	Patients with 2 or 3 events	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xxx(xx.x%)
	Patients with ≥4 events	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xxx(xx.x%)
Preferred term 1	Total patients	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
	Patients with 1 event	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xxx(xx.x%)
	Patients with 2 or 3 events	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xxx(xx.x%)
	Patients with ≥4 events	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xxx(xx.x%)
Preferred term 2	Total patients	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)	xxx (xx.x%)
	Patients with 1 event	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xxx(xx.x%)
	Patients with 2 or 3 events	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xxx(xx.x%)
	Patients with ≥4 events	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)	xxx(xx.x%)
...					

7.1.3.4. Time blocks

For certain events of interest, particularly common events in a long-term study, it may be important to understand how the risk of an event (e.g. assessed by percentage of patients, EAIR or EAER) or the risk of the drug changes (e.g. assessed by incidence ratio, incidence rate ratio, hazard ratio, rate or risk difference) over time. This point is mentioned in regulatory guidance documents such as a European guidance on anti-cancer products [51, Section "Temporal perspective"] and an FDA document [3, Section 7.4.1 "Common Adverse Events"]. As noted in Section 7.1.3.2, Kaplan-Meier curves are one way to investigate the time course of an event. Kaplan-Meier cumulative incidence can be used to estimate the percentage of patients who experienced an event up to a given time point. In this section, we focus on breaking the entire study period into smaller blocks of time in order to assess risk of an event occurring during a specific time period as opposed to cumulative incidence over time. Analysing by time blocks provides more information about the time course of an event than an analysis that covers the entire time period of the study. Keep in mind when interpreting results in one time block compared with another, ascertainment of AEs could be different, e.g. if the first 6 months on study has several visits and the second 6 months has few visits, the first 6 months might include more events than the second 6 months, merely because the more frequent data capture led to more events being reported.

When the time blocks are small and of equal duration, it may be sufficient to use percentage of patients (together with related metrics such as risk difference, risk ratio, odds ratio) as the analytical metric. In that case, a table (or figure) of percentage of patients who experience an event within a time block, such as Table 7.4, is useful. While Table 7.4 displays only one treatment arm, it is certainly possible to create a similar display with additional treatment arms. The percentage of patients could be calculated as follows:

- Each time interval could include all patients who enter the interval and could count the number of patients who experienced an incident event within the time interval (regardless of what happened in the prior interval). This

would give an idea of the magnitude of incidence across time intervals. It does not discriminate between new patients experiencing the event and patients who have a recurrent event.

- Each interval could include all patients who enter the interval, except those who previously experienced the event. This analysis can be used to easily identify when the first events per participant are occurring. It can answer a question like "if a participant has not had an event early on, how likely are they to have it later?"

The benefits of analysing the percentage of patients with an event are that it is somewhat easy to program and to understand (as long as the table is clear about whether the display is portraying situation a. or b. above). It gives more information about the time course of an event than a standard TEAE table that covers the entire time period of the study. However, this kind of analysis requires all of the time blocks to be the same duration because comparing percentage of patients with an event over time blocks of unequal duration is not sensible. If the time blocks are of differing durations, there is substantial differential dropout, or even if there is a desire to compare study incidence rates to literature-based incidence rates, a method that adjusts for person time at risk such as EAIR or EAER is useful. A table or figure of the EAIR could be produced within blocks of time. This could be in the form of a hazard curve (as shown in Figure 11 of Amit et al. [36]) or it could be like Figure 7.4. Though not displayed here, Figure 7.4 could be a split plot with the CIs for the differences included in a second panel. These figures provide a more direct assessment of the risk of an event occurring at any given time during the follow-up period, compared with a Kaplan-Meier curve. EAIR would only include the first event (ever) for each participant. If interest is in recurrent events, an appropriate metric is EAER. As noted in the PHUSE white paper on AEs [1], these metrics assume constant risk (constant hazard) over time; however, these assumptions are less problematic in smaller time blocks.

The decision of whether to analyse the first event per participant, or to analyse recurrent events will depend on the questions one wants to answer, the rarity of the event and its tendency to be episodic or not. Clearly, for non-episodic events, the first event would always be the one to analyse. Similarly, even for events that can be repeated, if the event is very rare, it may not be sensible to analyse recurrent events. For many events of interest, both types of analysis may be of interest. See Section 7.1.3.3 for more information on this topic.

Table 7.4. Example table that shows percentage of patients with an event in 90-day intervals. This could be either the first event per participant ever or the first event within an interval.

	----------INCIDENCE OF EVENTS WITHIN TIME INTERVALS----------				
	1-90	91-180	181-270	271-360	361-450
SYSTEM ORGAN CLASS	(N = XXX)	(N = XXX)	(N = XXX)	(N = XXX)	(N = XXX)
PREFERRED TERM	n (%)	n (%)	n (%)	n (%)	n (%)
ANY TEAE related to topic xxx	xxx (xx.x)	xxx (xx.x)	xxx (xx.x)	xxx (xx.x)	xxx (xx.x)
PT #1	xxx (xx.x)	xxx (xx.x)	xxx (xx.x)	xxx (xx.x)	xxx (xx.x)
PT #2	xxx (xx.x)	xxx (xx.x)	xxx (xx.x)	xxx (xx.x)	xxx (xx.x)
. . .	xxx (xx.x)	xxx (xx.x)	xxx (xx.x)	xxx (xx.x)	xxx (xx.x)

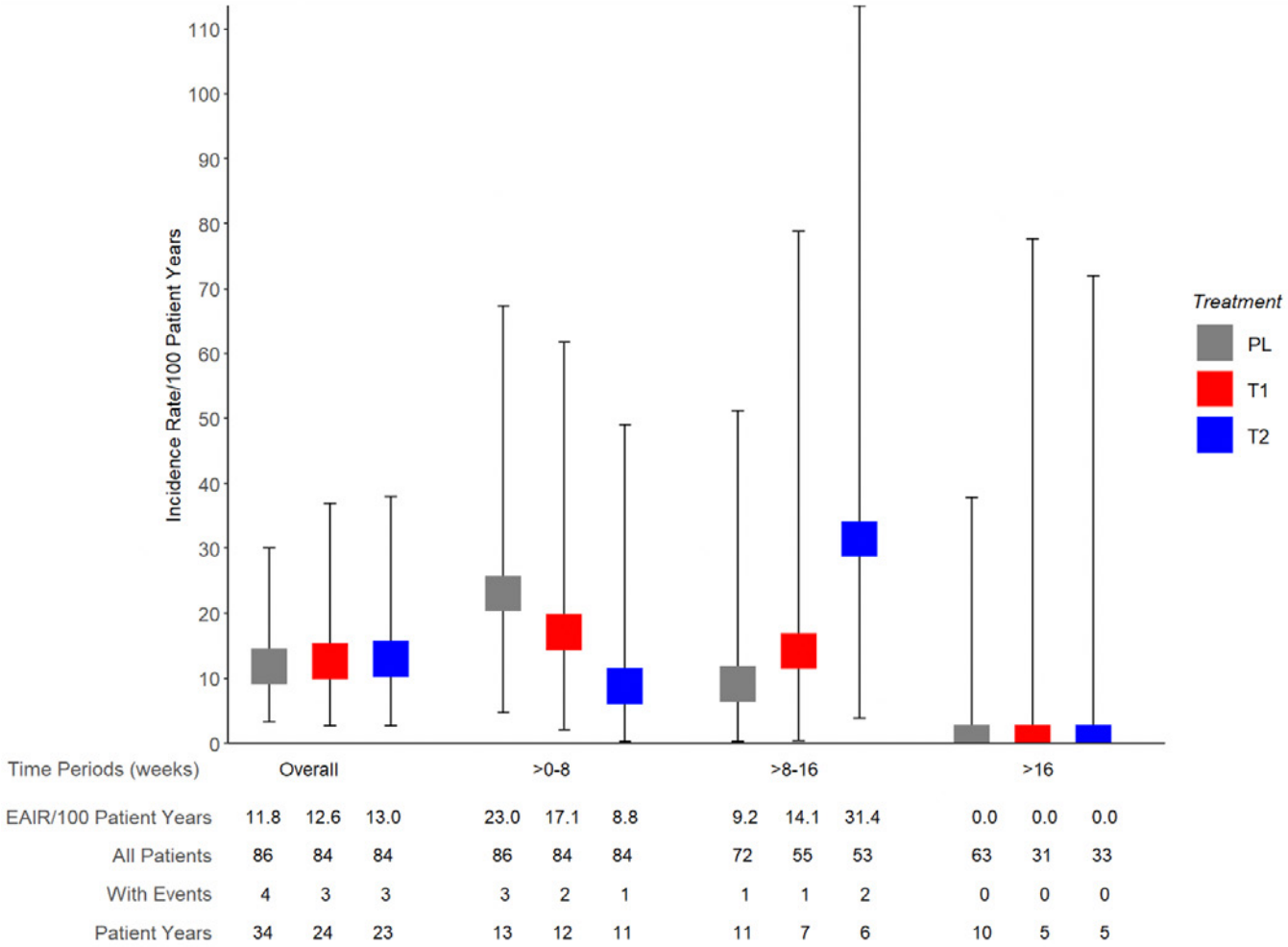


Figure 7.4. Example of EAIR in Time Blocks with 95% Confidence Intervals

7.1.4. Adverse Events of Special Interest that Include a Group of MedDRA preferred terms

The MedDRA dictionary is organised by a hierarchy with system organ class (SOC) at the top, followed by high level group terms (HLGTs), high level terms (HLTs), preferred terms (PTs) and, finally, lowest level terms (LLTs). (See Figure 7.5.) Each PT of MedDRA is assigned one primary SOC; all other SOC assignments for that PT are called "secondary". The MedDRA dictionary lists preferred terms in which the SOC/HLT are either the primary or a secondary SOC/HLT. In most cases, the study database and compound's integrated database only populates the primary SOC/HLT. So, if a team looks for a flag associated with a particular SOC/HLT, only those preferred terms in which the SOC/HLT is the primary mapping get picked up. It is possible that only the preferred terms that have a primary mapping to the SOC/HLT are the only ones of interest. However, it is also possible that all the preferred terms are of interest (primary and secondary). It is important that analytical plans are clear so that programming can be conducted in a manner to get the desired result. Furthermore, when sponsors create a custom set of preferred terms (PTs), also known as a Custom MedDRA Query (CMQ), the sponsor should provide the PTs used for the CMQ in a dataset. It is good practice to have the regulatory approving body review the list of PTs to ensure that they are appropriate. This streamlines the process of loading the list of PTs into analysis tools.

Some examples of MedDRA term assignments are as follows:

Congenital disorder: PT congenital absence of bile ducts
- Primary SOC: Congenital, familial and genetic disorders
- Secondary SOC: Hepatobiliary disorders
  - Secondary SOC assignment is based on the site of manifestation

Neoplastic disorder: PT skin cancer
- Primary SOC: Neoplasms benign, malignant and unspecified (including cysts and polyps)
  - Primary SOC assignment for cyst and polyp is the site of manifestation
- Secondary SOC: Skin and subcutaneous tissue disorders

Infectious disorder: PT: Enterocolitis infectious ducts
- Primary SOC: Infections and infestations
- Secondary SOC: Gastrointestinal disorders

In addition to this organisation, the MedDRA dictionary includes standardised MedDRA queries (SMQs). SMQs are groupings of terms, usually PTs, that relate to a defined medical condition or area of interest [52]. If a standardised MedDRA query (SMQ) exists for a particular AESI, it is recommended to analyse the data using the SMQ. The main advantages of the SMQs are that they 1) have been developed by an external medical group and do not have the appearance of bias in their creation, 2) are

updated with each version of MedDRA, and 3) are recognised by regulatory agencies. Because newer MedDRA versions have more SMQs, studies that used old versions of MedDRA should be recoded to a newer version of MedDRA for a more comprehensive integrated safety review.

It is important to review the list of PTs included in an SMQ. Some SMQs exclude PTs that are presumed to not be drug-related. For example, the SMQ Acute Pancreatitis excludes the PT's "alcoholic pancreatitis", "autoimmune pancreatitis", "obstructive pancreatitis", and infectious causes such as "pancreatitis viral". There are several reasons why these exclusions may be problematic: 1) the etiology specified by the PT may represent the mechanism by which a medication causes an adverse reaction, 2) patients experiencing the AE might be more susceptible to a medication that causes the adverse reaction, and if so, it might be prudent to identify that risk group, and 3) clinical trial investigators may inaccurately attribute the cause of an adverse event (e.g. pancreatitis in a patient with a history of alcoholic pancreatitis might lead an investigator to incorrectly attribute the event to alcoholic pancreatitis).

For special topics for which an SMQ does not exist, sometimes MedDRA HLGTs (such as gastrointestinal signs and symptoms) and/or MedDRA HLTs (such as nausea and vomiting) or a CMQ are useful.

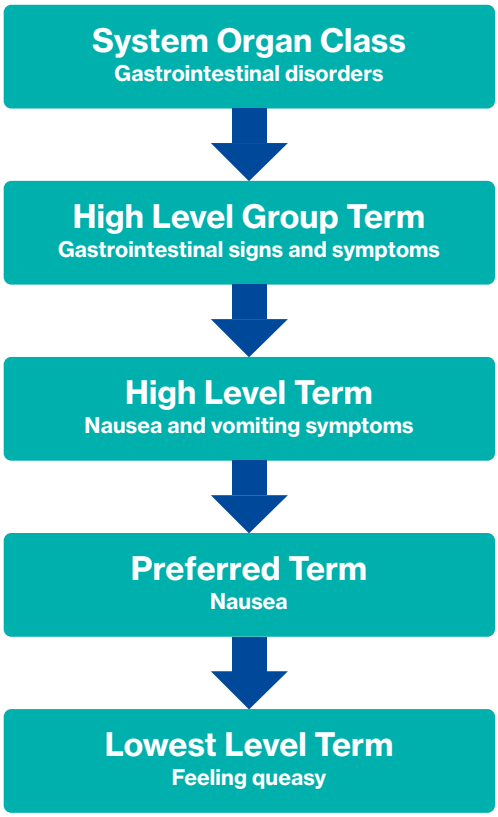


Figure 7.5. MedDRA Hierarchy of Descriptive Terms: Example of How One Term Maps in the Hierarchy

In addition to SMQs, MedDRA hierarchical groupings, and industry internal custom lists, the FDA has an effort to develop standard queries for detecting and summarising safety signals called the FDA MedDRA query (FMQ) project [37]. The FDA examined more than 38,000 labels using natural language processing and is developing queries using preferred terms for the most frequently labelled adverse events.

Sometimes an endpoint consists of a composite of multiple events, such as MACE (a major adverse cardiovascular event, which often includes non-fatal myocardial infarction, non-fatal stroke and cardiovascular death). If part of a planned statistical analysis, the components of the composite are typically displayed in a table underneath the composite. If the risk of each component is in the same direction/consistent with the composite, then the overall results are more reassuring. For example, if a component included death, but the risk of mortality was in the opposite direction, this would raise questions/doubt of the validity of including mortality in the claim.

Organisations should specify as early as possible their planned analyses for AESIs, including listing the preferred terms that will be used in the clinical study protocol or the statistical analysis plan. Having this clarity in advance improves collection of events, programming planning and data analysis. An example of how to be clearer on the intent in analysis plans: Instead of hepatitis will be defined as TEAEs from the MedDRA HLT "Hepatic viral infections" or the HLT "Hepatic infections (excl viral)", it would be helpful to also list the PTs under those two HLTs.

The following tables are useful for various situations. Table 7.5 is used for any predefined cluster of PTs that do not have any subclassification. Table 7.6 is useful for a simple SMQ that has broad- and narrow-scope terms. Table 7.7 is useful for an SMQ that has sub-SMQs as well as broad- and narrow-scope terms. Of note, when an SMQ is used, including counts and percentages of patients reporting, at least one PT in the broad- and narrow-scope list has the potential for misinterpretation. Since broad terms are used to cast a large net of events potentially related to the topic at hand, any count and percentage provided that includes these terms would likely be an over-estimation within study-arm and simultaneously yield a dilution of treatment effect. For additional information, please see the section "Conservativeness" in the AE white paper [1]

Table 7.5. Summary of [Adverse Events of Special Interest] Using a Pre-Defined List of MedDRA Preferred Terms by Preferred Term in Descending Frequency of T1 & T2

Preferred Term	PL (N=XX) n (%)	T1 (N=XX) n (%)	T2 (N=XX) n (%)	T1 & T2 (N=XX) n (%)	Risk Difference[a] T1-PL (95% CI)	Risk Difference[a] T2-PL (95% CI)	Risk Difference[a] T1&T2-PL (95% CI)
Number of patients reporting at least one [AESI] adverse event	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #1]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #2]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #3]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #n]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)

Footnotes:
N = number of patients; n = number of patients with at least one row event. Percentages are calculated relative to the treatment group N.
Patients may be counted in more than one row. T1 = Treatment 1; T2 = Treatment 2; PL = Placebo.
[a]Risk difference confidence interval from [insert method].

Table 7.6. Summary of [Adverse Events of Special Interest] Defined Using Standardised MedDRA Query [SMQ Name]

Preferred Term	PL (N=XX) n (%)	T1 (N=XX) n (%)	T2 (N=XX) n (%)	T1 & T2 (N=XX) n (%)	Risk Difference[a] T1-PL (95% CI)	Risk Difference[a] T2-PL (95% CI)	Risk Difference[a] T1&T2-PL (95% CI)
Number of patients reporting at least one Narrow-scope PT	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
Narrow-scope PTs							
[Preferred Term #1]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #2]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #n]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
Broad-scope PTs							
[Preferred Term #1]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #2]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #n]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)

Footnotes:
Narrow- and broad-scope preferred terms (PTs) are defined by a standardised MedDRA SMQ [or, if not by MedDRA, provide appropriate detail, e.g. sponsor-defined custom search]. T1 = Treatment 1; T2 = Treatment 2; PL = Placebo.
N = number of patients; n = number of patients with at least one row event. Percentages are calculated relative to the treatment group N.
Patients may be counted in more than one row.
[a]Risk difference confidence interval from [insert method].

Table 7.7. Summary of [Adverse Events of Special Interest] Defined Using [SMQ Name – SMQ with Sub-SMQs]

Query and Scope Preferred Term	PL (N=XX) n (%)	T1 (N=XX) n (%)	T2 (N=XX) n (%)	T1 & T2 (N=XX) n (%)	Risk Difference^ T1-PL (95% CI)	Risk Difference^ T2-PL (95% CI)	Risk Difference^ T1&T2-PL (95% CI)
Patients with at least one narrow-scope PT							
{ SMQ Name }	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
sub-SMQ1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
sub-sub-SMQ1.1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
sub-sub-SMQ1.2	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
sub-SMQ2	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
sub-sub-SMQ2.1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
...							
Narrow-scope PTs							
{ SMQ Name }							
sub-SMQ1							
sub-sub-SMQ1.1							
[Preferred Term #1]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #2]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
sub-sub-SMQ1.2					xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
sub-SMQ2							
sub-sub-SMQ2.1							
[Preferred Term #]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
...							
Broad-scope PTs							
{ SMQ Name }							
sub-SMQ1							
sub-sub-SMQ1.1							
[Preferred Term #1]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #2]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
sub-sub-SMQ1.2							
[Preferred Term #]	xx (xx.x)				xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
sub-SMQ2							
sub-sub-SMQ2.1							
[Preferred Term #]	xx (xx.x)				xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
. . .							

Abbreviations: PBO = Placebo; PT = Preferred term; TRT = Treatment.
Footnotes:
Narrow- and broad-scope preferred terms (PTs) are defined by a standardised MedDRA SMQ.
N = number of patients; n = number of patients with at least one row event. Percentages are calculated relative to the treatment group N.
Patients may be counted in more than one row.
^Risk difference confidence interval from [insert method].

7.2. Discussion

This paper contains suggested analyses and output layouts to address common safety topics of interest. Many of these analyses and data displays can be used directly or with small modifications to address additional safety topics. Likewise, the examples provided are not exhaustive; there are many useful analyses that are not included here that may be applied to the safety topics of interest covered in this paper.

One topic that was discussed, as the example shells were created, was the order of the treatment groups. If treatment groups were ordered with placebo first (e.g. placebo, low dose, high dose), it would be easier to see a potential dose relationship. However, others prefer the investigational product to appear first. In this white paper, we chose the ordering with placebo first. However, compound teams will likely develop conventions, and it's best to follow the same convention throughout a submission.

8. Tables and Figures for Integrated Summaries

8.1. Recommended Displays

The recommended displays for integrated summaries are essentially the same as for individual studies. However, the methodology for any summary metric, confidence interval and/or p-value will likely need to incorporate stratification by study (see Section 6.4). For our sample displays, we assume not all studies will have the same doses, and that all doses of the study drug that fall within the range of draft label dosing will be included as a single treatment arm. Thus, unlike the sample displays for individual studies, the sample displays for integrated summaries will have a single treatment arm and a placebo arm. We will also assume (after pooling) that the treatment-placebo randomisation ratio is not the same across the studies, so the study-size adjusted percentages are included in the example displays. The tables can be modified as needed under different scenarios. The examples also include 95% confidence intervals.

As noted in Section 10.1 of PHUSE 2017 [1], if a sponsor decides not to utilise confidence intervals as a tool for review, they can be deleted. In addition, if a sponsor decides to utilise p-values as a tool for review, they can be added.

There is one recommended table for integrated summaries that is a bit different from what is provided for individual studies. This assumes there is uncontrolled data in the clinical program (e.g. extension periods or extension studies). See Table 8.1. This table provides EAIRs for a safety population including all time on drug for patients with any time on drug ("All Drug", controlled and uncontrolled). EAIRs for an integrated controlled analysis set are included to provide context. As noted in the recommended footnote, comparing these EAIRs is problematic. If the EAIR is similar between the All Drug group and the placebo group, it cannot be used as evidence towards a lack of difference between treatment and placebo. The constant rate assumption is almost surely in violation. However, if the EAIR is greater in the All Drug group than the placebo group, that event might warrant further scrutiny.

Additionally, the recommended figure in Section 7.1.2 for events with very low frequency (Figure 7.2) may not be reasonable for an integrated summary if there are more than a few studies in the clinical programme. See Table 8.2 for a recommendation in these cases.

Table 8.1. Person-Time-Adjusted Incidence Rates for All Drug, All Time and Placebo-controlled Periods

Preferred Term	Placebo during placebo-controlled period all indications (N=xxx, PY=xxx) n, PYE, IR; 95% CI)	DRUG during placebo-controlled period all indications (N=xxx, PY=xxx) n, PYE, IR; 95% CI)	All DRUG, all time n, PYE, IR; 95% CI)
Preferred Term 1	xx, XXX, (xx.x; xx.x, xx.x)	xx, XXX, (xx.x; xx.x, xx.x)	xx, XXX, (xx.x; xx.x, xx.x)
. . .	…	…	
Preferred Term n	xx, XXX, (xx.x; xx.x, xx.x)	xx, XXX, (xx.x; xx.x, xx.x)	xx, XXX, (xx.x; xx.x, xx.x))

Abbreviations: CI = confidence interval; IR = incidence rate; N = number of patients; n = number of patients with at least one treatment-emergent adverse event; PY = total patient years; PYE = patient years at risk.
IR is 100 times the number of patients experiencing the adverse event divided by the event-specific exposure to treatment (exposure time up to the event for patients with the event and exposure time to the end of the period for patients without the event, in years).
Note: Summary statistics from the placebo-controlled period is provided for context. Comparing All DRUG to other results is problematic. Study and treatment are confounded, and risk over time can change due to reasons other than treatment exposure. Interpretation is challenging and has similar limitations to observational data.

Table 8.2. Collage of Incidence of Treatment-Emergent [Event Cluster]

Study Event Term, n (%)	Placebo *a	Active Comparator *b	TrtDose1	TrtDose2	TrtDose3	TrtDose4	TrtDose5
Study 1 (XXXX)	N=xx				N=xx		N=xx
No Events							
Study 2 (XXXX)	N=xx		N=xx		N=xx	N=xx	N=xx
No Events							
Study 3 (XXXX)		drug name N=xx		N=xx	N=xx		
No Events							
Study 4 (XXXX)	N=xxx			N=xxx		N=xxx	
No Events							
Study 5 (XXXX)		drug name N=xxx		N=xxx		N=xxx	
PT name 1		0 (0.0)		2 (0.7)		0 (0.0)	
Study 6 (XXXX)	N=177	drug name N=439*b N1=315, N2=124	N=49	N=302	N=10	N=304	N=45
PT name 2	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.3)	1 (2.2)
PT name 3	0 (0.0)	1 (0.2)	0 (0.0)	1 (0.3)	0 (0.0)	1 (0.3)	0 (0.0)

Abbreviations: N = patients assigned to this treatment group who received at least one dose of study medication; N1 = patients who received this treatment starting at randomisation; N2 = patients who received this treatment after 6 months of placebo; n = number of patients in specified category.
*a - The planned Placebo period in studies X, Y, Z is 26 weeks, which is shorter than the overall study durations of 104 weeks, 52 weeks and 26 weeks, respectively. Rates in the Placebo group are not directly comparable to rates in other groups.
*b - Active comparator and treatment include any patients from Studies X and Y, respectively, who began the study receiving placebo and reached the timepoint where they began to receive active treatment (indicated by N2). If an event is treatment-emergent relative to original baseline after that timepoint, it is included in this table (independent of the Placebo period).

9. Disclaimer

The opinions expressed in this document are those of the contributors and should not be construed to represent the opinions of PHUSE (Pharmaceutical User Software Exchange), members' respective companies or organisations, or the FDA's views or policies. The content in this document should not be interpreted as a data standard and/or information required by regulatory authorities.

10. Revision History

Date	Author	Version	Changes
Overall	See Section 11	v1.0	First edition

11. Acknowledgements

The primary contributors include Brenda Crowe, Mary Nilsson, Aimee Cyr, Bach Nhi Beasley, Joan Buenconsejo, Swarna Reddy, Wei Wang, Monali Desai and Rebeka Revis.

Additional contributors/reviewers (e.g. participated in discussions, provided review comments on portions of the content) include Nancy Brucken, Jim Cawley, William Holmes, Bill Palo, Adrienne Rothstein, Norman L Stockbridge, and Aiming Yang.

12. Project Contact Information

Email: *workinggroups@phuse.global*

13. References

All links in this section were accessed 21st January, 2021.

1. PHUSE (2017). Analysis and displays associated with adverse events: Focus on adverse events in phase 2-4 clinical trials and integrated summary documents.

2. US Food and Drug Administration. (2005). Reviewer guidance: Conducting a clinical safety review on a new product application and preparing a report on the review. [serial online]. Available from: *https://www.fda.gov/media/71665/download*

3. US Food and Drug Administration. (2010). Good review practice: Clinical review template. *Manual of policies and procedures* [serial online]. Available from: *https://www.fda.gov/media/72472/download*

4. CIOMS Working Group VI. (2005). Management of safety information from clinical trials. Geneva: Council for International Organizations of Medical Sciences.

5. Ma, H., Ke, C., Jiang, Q., et al. (2015) Statistical considerations on the evaluation of imbalances of adverse events in randomized clinical trials. *Therapeutic Innovation & Regulatory Science, 49*, 957–965.

6. PHUSE. (2013). Analyses & Displays Associated with Measures of Central Tendency - Focus on Vital Sign, Electrocardiogram, & Laboratory Analyte Measurements in Phase 2-4 Clinical Trials & Integrated Submission Documents. [serial online]. Available from: *https://www.phuse.eu/documents//working-groups/deliverables/analyses-displays-associated-with-measures-of-central-tendancy-version-10-10-oct-13-11816.pdf*

7. PHUSE. (2015). Analyses and Displays Associated with Outliers or Shifts from Normal to Abnormal: Focus on Vital Signs, Electrocardiogram, and Laboratory Analyte Measurements in Phase 2-4 Clinical Trials and Integrated Summary Documents. [serial online]. Available from: *https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Standard+Analyses+and+Code+Sharing/Analyses+%26+Displays+Associated+with+Outliers+or+Shifts+from+Normal+To+Abnormal+Focus+on+Vital+Signes+%26+Electrocardiogram+%26+Laboratory+Analyte+Measurements+in+Phase+2-4+Clinical+Trials+and+Integrated+Summary.pdf*

8. US Food and Drug Administration. (2012). *Guidance for industry and investigators: Safety reporting requirements for INDs and BA/BE studies.* Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research.

9. Zhou, Y., Ke, C., Jiang, Q., et al. (2015). Choosing appropriate metrics to evaluate adverse events in safety evaluation. *Therapeutic Innovation & Regulatory Science.* 2168479014565470.

10. Rothman, K., Greenland, S., & Lash, T. (2008). Modern epidemiology. Philadelphia, PA: Lippincott, Williams & Wilkins.

11. (2005). *Handbook of epidemiology.* Berlin, Heidelberg, New York: Springer-Verlag Berlin Heidelberg.

12. Chuang-Stein, C., & Beltangady, M. (2011). Reporting cumulative proportion of subjects with an adverse event based on data from multiple studies. *Pharm Stat, 10,* 3–7.

13. Crowe, B., Chuang-Stein, C., Lettis, S., et al. (2016). Reporting adverse drug reactions in product labels. *Therapeutic Innovation & Regulatory Science, 50,* 455–463.

14. Rosario, L. A. , Kropp, T. J., Wilson, S. E., et al. (2012). Join FDA/PhUSE working groups to help harness the power of computational science. *Drug Inf J, 46,* 523–524.

15. US Food and Drug Administration. (2016). Guidance for Industry: Determining the extent of safety data collection needed in late-stage premarket and postapproval clinical investigations. [serial online]. Available from: *https://www.fda.gov/media/82664/download*

16. CDISC Analysis Data Model Team. (2012). Analysis data model (ADaM) data structure for adverse event analysis. [serial online]. Available from: *http://clinicalsafetygeek.com/wp-content/uploads/2013/08/adam_ae_final_v1.pdf*

17. CIOMS Working Group X. (2016). Evidence synthesis and meta-analysis for drug safety: Report of the CIOMS working group X. Geneva, Switzerland: Council for International Organizations of Medical Sciences (CIOMS).

18. US Food and Drug Administration. (2013). Guidance for industry: E3 structure and content of clinical study reports. *Questions and answers (R1)* [serial online]. Available from: *http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm336889.pdf*

19. Japanese Administrative Notice. (2011). Format for preparing the common technical document for submission of New Drug Applications to reduce total review time. [serial online]. Available from: *http://home.att.ne.jp/red/akihiro/mhlw/CTD_20110322.pdf*

20. European Medicines Agency (2014). Guidance document for the content of the <Co->Rapporteur day 80 critical assessment report. [serial online]. Available from: *http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2009/10/WC500004856.pdf*

21. International Conference on Harmonisation Expert Working Group. (1995). Structure and content of clinical study reports: E3. *ICH harmonised tripartite guideline* [serial online].

22. International Conference on Harmonisation Expert Working Group. (2002). The Common Technical Document for the registration of pharmaceuticals for human use: Efficacy M4E(R1). *Clinical Overview and Clinical Summary of Module 2 and Module 5: Clinical Study Report*s [serial online].

23. International Conference on Harmonisation. (1998). E9: Statistical Principles for Clinical Trials. *International Conference on Harmonisation Guidelines* [serial online]. Available from: *https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials*

24. US Food and Drug Administration. (2005). Guidance for Industry: Premarketing risk assessment. [serial online]. Available from: *https://www.fda.gov/regulatory-information/search-fda-guidance-documents/premarketing-risk-assessment*

25. European Commission; Enterprise and Industry Directorate-General. (2009). A guideline on Summary Of Product Characteristics (SmPC). [serial online]. Available from: *http://ec.europa.eu/health/files/eudralex/vol-2/c/smpc_guideline_rev2_en.pdf*

26. Schechtman, E. (2002). Odds ratio, relative risk, absolute risk reduction, and the number needed to treat—which of these should we use? *Value Health, 2*(5), 431–436.

27. Crowe, B., Xia, H. A., Berlin, J., et al. (2009). Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: A report of the safety planning, evaluation, and reporting team. *Clin Trials, 6,* 430–440.

28. International Conference on Harmonisation. ICH E9 statistical principles for clinical trials ICH Harmonised Tripartite Guideline. 1995 [25 May 2015]. Available from: *http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/statistical-principles-for-clinical-trials.html*

29. Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. *The American Statistician, 70,* 129–133.

30. Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a World Beyond "p<0.05". *The American Statistician, 73,* 1–19.

31. Talbot, J., & Aronson, J. K. (Eds.). (2012). *Stephens' detection and evaluation of adverse drug reactions: Principles and practice.* Chichester, West Sussex, UK: John Wiley & Sons, Ltd.

32. CIOMS Working Groups III and V. (1999). Guidelines for preparing core clinical safety information on drugs. Geneva: Council for International Organizations of Medical Sciences.

33. Hill, A. B. (1965). The environment and disease: Association or causation? *Proc R Soc Med, 58,* 295–300.

34. Fedak, K. M., Bernal, A., Capshaw, Z. A., et al. (2015). Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. *Emerging themes in epidemiology, 12*(14).

35. Duke, S. P., Bancken, F., Crowe, B., et al. (2015). Seeing is believing: Good graphic design principles for medical research. *Stat Med.*

36. Amit, O., Heiberger, R., & Lane, P. (2008). Graphical approaches to the analysis of safety data from clinical trials. *Pharm Stat, 7,* 20–35.

37. Sutter, S. (September 2019). US FDA Looks To Standardize Premarketing Safety Assessments. *Pink Sheet* [serial online]*, 8.*

38. Unkel, S., Amiri, M., Benda, N., et al. (2019). On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. *Pharm Stat, 18,* 166–183.

39. Hengelbrock, J., Gillhaus, J., Kloss, S., et al. (2016). Safety data from randomized controlled trials: applying models for recurrent events. *Pharm Stat, 15,* 315–323.

40. Proctor, T., & Schumacher, M. (2016). Analysing adverse events by time-to-event models: the CLEOPATRA study. *Pharm Stat, 15,* 306–314.

41. Allignol, A., Beyersmann, J., & Schmoor, C. (2016). Statistical issues in the analysis of adverse events in time-to-event data. *Pharm Stat, 15,* 297–305.

42. Bender, R., Beckmann, L., & Lange, S. (2016). Biometrical issues in the analysis of adverse events within the benefit assessment of drugs. *Pharm Stat, 15,* 292–296.

43. Pintilie, M. (2006). *Competing Risks: A Practical Perspective.* John Wiley & Sons, Ltd.

44. Allison, P. D. (2018). For Causal Analysis of Competing Risks, Don't Use Fine & Gray's Subdistribution Method. [serial online]. Available from: *https://statisticalhorizons.com/for-causal-analysis-of-competing-risks*

45. PHUSE 2019. Sharing an Interactive Display in an FDA Submission Pilot [Webinar Slides]. Available from: *https://phuse.s3.eu-central-1.amazonaws.com/Advance/Data+Visualisation+and+Open+Source+Technology/Interactive+Data+Visualizations+for+Decision+Making+in+Submissions+%E2%80%93%C2%A0PHUSE+Webinar.pdf*

46. PHUSE 2019. Sharing an Interactive Display in an FDA Submission Pilot [Webinar Recording]. Available from: *https://www.youtube.com/watch?v=u26-A2NTj2k*

47. Rubin, D. B. (2011). An alternative to pooling Kaplan-Meier curves in time-to-event meta-analysis. *The International Journal of Biostatistics, 7.*

48. Kemmler, G., Hummer, M., Widschwendter, C., et al. (2005). Dropout rates in placebo-controlled and active-control clinical trials of antipsychotic drugs: a meta-analysis. *Arch Gen Psychiatry, 62,* 1305–1312.

49. Cook, R. J., & Lawless, J. F. (2007). *The statistical analysis of recurrent events.* New York, NY: Springer.

50. Reliawiki. *Recurrent event data analysis/Mean cumulative function.* [serial online]. Available from: *http://reliawiki.org/index.php/Recurrent_Event_Data_Analysis#:~:text=The%20Mean%20Cumulative%20Function%20(MCF),-In%20a%20non&text=It%20is%20a%20staircase%20function,a%20unit's%20cumulative%20history%20function*

51. European Medicines Agency. (2017). Guideline on the evaluation of anticancer medicinal products in man. [serial online]. Available from: *https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-evaluation-anticancer-medicinal-products-man-revision-5_en.pdf*

52. MedDRA. Standardised MedDRA Queries (SMQs). [serial online]. Available from: *https://www.meddra.org/how-to-use/tools/smqs*