Analysis and Displays Associated with Adverse Events: Focus on Adverse Events in Phase 2–4 Clinical Trials and Integrated Summary Documents

phuse.eu

Contents

1. Disclaimer

The opinions expressed in this document are those of the authors and should not be construed to represent the opinions of the Pharmaceuticals User Software Exchange (PHUSE), members' respective companies or organizations, or 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.

2. Notice of Current Edition

This edition of the "Analysis and Displays Associated with Adverse Events: Focus on Adverse Event in Phase 2-4 Clinical Trials and Integrated Summary Documents" is the first edition.

3. Additions and/or Revisions

Date	Author	Version	Changes
2017-FEB-03	See Section 14	v1.0	Final edition

4. Overview: Purpose

This white paper provides advice on displaying, summarizing, and analyzing adverse events (AEs) in tables, figures, and listings (TFLs), with a focus on Phase 2-4 clinical trials and integrated summary documents. The intent is to begin the process of developing a common standard for analysis and reporting of collected information that is used across clinical trials and across therapeutic areas. Although the white paper focuses on displays and analyses of data already collected, it will have implications for what and how data should be collected.

The development of standard TFLs and associated analyses will lead to improved standardization of AE data from collection through data display and analysis. The development of standard TFLs will also lead to improved and harmonized product lifecycle management across therapeutic areas by ensuring that reviewers receive clinically relevant and meaningful analyses of patient safety for benefit-risk assessment. 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 to understand and reduce bias in analysis planning and reporting. This assistance is important even when inferential statistics are not used. The potential for biased comparisons is especially a concern when multiple studies are combined (eg, via poor pooling practices for integrated summaries). Safety physicians have 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 [1], while medical judgment 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 expertise. For example, all members of a safety management team should understand that improper pooling can lead to biased and even paradoxical results. See Table 10.1 for an illustrative example of paradoxical results and see Section 10.1 for other areas where increased expertise is needed.

5. Scope

This white paper is intended to provide advice to sponsors who are developing the analysis plan for Phase 2-4 clinical trials and integrated summary documents (or other documents in which analysis of AEs are of interest). The advice pertains to AE analyses that are generally relevant across all therapeutic areas (treatment-emergent AEs [TEAEs], serious AEs [SAEs], AEs leading to discontinuation of study drug, deaths, etc). The advice also pertains to general safety signal detection for AEs. More specialized analyses or additional analyses may be needed for adverse events of special interest (AESIs), including events that become of interest after looking at the results from the initial set of analyses. General recommendations for analyses and displays that might be useful as the next step in further understanding and characterizing events are a topic for a future white paper.

Although the focus of this white paper primarily pertains to Phase 2-4 clinical trials, some of the content may apply to Phase 1 studies or other types of medical research (eg, observational studies).

Detailed variable specifications for TFLs or dataset development are out of scope. The PHUSE Repository Content and Delivery Project Team will be developing code (utilizing Study Data Tabulation Model [SDTM] and Analysis Data Model [ADaM] data structures from the Clinical Data Interchange Standards Consortium [CDISC]) that are consistent with the concepts outlined in this white paper and placed in the publicly available PHUSE Standard Scripts Repository.

6. Definitions

Term	Definition
ADaM	Analysis Data Model
AE	adverse event
AESI	adverse event of special interest
ADR	adverse drug reaction
CDASH	Clinical Data Acquisition Standards Harmonization
CDISC	Clinical Data Interchange Standards Consortium
CIOMS	Council for International Organizations of Medical Sciences
CS	computational science
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
EAIR	exposure-adjusted incidence rate
ECG	electrocardiogram
FDA	Food and Drug Administration
HLGT	high level group term
HLT	high level term
ICH	International Conference on Harmonization
IR	incidence rate
IRR	incidence rate ratio
LLIC	Linear Logistic Indirect Comparison model
LLT	lowest lever term
MEAD	MedDRA-Based Adverse Event Diagnostics
MedDRA	Medical Dictionary for Regulatory Activities
MSSO	Maintenance and Support Services Organization
NME	new molecular entity
PHUSE	Pharmaceuticals User Software Exchange
PMDA	Pharmaceuticals and Medical Devices Agency
PSAP	Program Safety Analysis Plan
PT	preferred term
SAE	serious adverse event
SAP	statistical analysis plan
SDTM	Study Data Tabulation Model
SOC	system organ class
SmPC	Summary of Product Characteristics
SMQ	standardized MedDRA query
TEAE	treatment-emergent adverse event
TFL	tables, figures, and listings
TRT	treatment
WHO	World Health Organization

Definitions

The following provides definitions for terms used in this white paper. In some cases, 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. For example, the number of patients with an event divided by the number of patients at risk for the event has been referred to as the event rate [2], incidence rate [3], crude incidence rate [4], incidence proportion [2, 5-7], cumulative incidence [2, 6, 7], or simply the percentage. Incidence rate has been used for the number of patients with an event divided by the number of patients at risk for the event [3] and for the number of patients with an event divided by time at risk for the event [2, 5-7]. Thus, it is important to include clear definitions in statistical analysis plans (SAP), clinical study reports (CSR), submission documents, and tables/figures (eg, 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 it should be clear whether the denominator is the number of patients in a population, the time at risk, or the total time exposed. We concur with sources that include "person-time" or "exposure-adjusted" in the name when time is used in the denominator for clarity.

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 (Section 1.2. of Attachment B of the Food and Drug Administration (FDA) Clinical Review Template [8]; See also Section 10.2 of this document).

Adverse event of special interest (AESI): An AE or group of AEs of scientific and medical concern specific to the sponsor's product or program, which may require further investigation in order to characterize and understand them [1]. Example AESIs include hepatic-related events, infections, and hypersensitivities. Some AESIs can be addressed using a Medical Dictionary for Regulatory Activities (MedDRA) standardized MedDRA query (SMQ) and should be addressed with a MedDRA SMQ if one exists for the medical concept of interest.

Adverse drug reaction (ADR): An undesirable effect, reasonably likely to be caused by a study drug and it may occur as part of the pharmacological action of the study drug or may be unpredictable in its occurrence. Section 1.2. of Attachment B of the FDA Clinical Review Template [8] and in the FDA Guidance on Safety Reporting Requirements for INDs and BA/BE Studies [9]. See also Section 10.2.

Consolidated terms: A list of terms (usually MedDRA preferred terms [PTs]) that are grouped together to address a particular medical concept of interest (eg, an AESI) or are grouped because they represent the same phenomenon (eg, sedation, somnolence, drowsiness). This is sometimes referred to as clustered events [10].

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. As discussed above, this is sometimes referred to as the event rate, incidence rate, crude incidence rate, incidence proportion, or cumulative incidence. This is the primary quantity used for signal detection and labeling in most cases (see Section 10.9).

Exposure-adjusted incidence rate (EAIR): The number of patients with an event divided by the total time at risk for the event [4]. Time at risk will be calculated as the sum of time to first event for patients who experienced the event and the time during the interval for patients who do not experience the event. This is sometimes referred to as the incidence rate or person-time incidence rate [6, 7]. As noted above, we believe the addition of "exposure-adjusted" or "person-time" is beneficial for clarity. The use of the EAIR is generally more appropriate for AESIs or for signal detection when there are differential dropouts between treatments, but would need to be used with caution (see Section 10.9).

Program safety analysis plan (PSAP): A compound-level planning document describing the planned analyses and definitions required to conduct the planned analyses for Phase 2-3 studies and the Summary of Clinical Safety for a compound [11]. The PSAP may also contain compound-level data collection requirements.

Serious adverse event (SAE): An AE, whether or not it is considered study-drug related that results in any one of a set of predefined criteria. Regulatory agencies define a set of minimum criteria (Section 1.2 of Attachment B of the FDA Clinical Review Template [8]; See also Section 10.2).

Static display: A table, figure, or listing that is created without any interactive features. It is something that can be viewed in its entirety on a page(s). 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 proportion. See Section 10.1 for an example.

Treatment-emergent adverse event (TEAE): An event not present at baseline, or not present at the severity seen on treatment (Section 7.4.1 of Attachment B of the FDA Clinical Review Template [8]). While this definition appears straightforward, there are multiples ways in which this is implemented across the industry (see Section 10.2).

7. 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 analyze safety signals from randomized clinical trials [14]. Several regulatory guidance documents contain suggested TFLs and/or discussion around analyses for common assessments and quantitative measurements, but they are open to varied interpretation. While individual companies may have their own standard TFLs for common data elements, lack of cross-industry standards requires medical reviewers to understand and learn the various nuances and methods across the submissions he/she reviews. In addition, if definitions are different or if the data are displayed in a manner that a medical reviewer is not accustomed to, the sponsor may be asked to summarize the data differently potentially leading to not meeting a first cycle review. Having many variations for defining certain terms, variations on how to include/exclude TEAEs occurring after stopping study drug, and variations in how the usual safety data are displayed (eg, AEs, SAEs, etc) leads to unnecessary burden. Cross-industry standardization also provides a mechanism to incorporate shared learning more efficiently and potentially enables quicker implementation of new methodologies. Crowd-sourcing standardization efforts (instead of each sponsor networking to maintain their own standards) should lead to more optimal displays for medical reviewers, health authorities, ethics committees, and drug development teams. While individual companies with mature standardization in place may be hesitant to change to the recommendations in this white paper, the idea is that, over time, the overall review of safety data will be improved by implementing the latest recommendations and by gaining familiarity of definitions, methods, and displays among all those involved. Using the recommendations in this white paper should lead to more effective and clear communication for all stakeholders.

8. Background

The PHUSE Computational Science Collaboration is an initiative between PHUSE, the FDA, and industry that identified computational science priorities that could be addressed by collaboration, crowd sourcing, and innovation [15]. Several computational science (CS) working groups were created to address many of these challenges. The working group titled "Standard Analyses and Code Sharing" (formerly "Development of Standard Scripts for Analysis and Programming") 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 (see bulleted list below) 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. The documents include the following:

1. Council for International Organizations of Medical Sciences (CIOMS) X – Evidence Synthesis and Meta-Analysis for Drug Safety (2016) [1].
2. FDA Guidance for Industry: Collection Needed in Late Stage Premarket and Postapproval Clinical Investigations (2016) [16].
3. European (EU) Guidance document for the content of the <Co-> critical assessment report (2014) [17].
4. FDA Guidance for Industry: E3 Structure and Content of Clinical Study Reports Questions and Answers (R1) (2013) [18].
5. FDA Guidance for Industry and Investigators: Safety Reporting Requirements for INDs and BA/BE Studies (2012) [9].
6. ADaM Data Structure for Adverse Event Analysis (2012) [19].
7. Japan Format for Preparing the Common Technical Document for Submission of New Drug Applications to Reduce Total Review Time (2011) [20].
8. FDA Manual of Policies and Procedures: Clinical Review Template (2010) [Attachment B contains the detail related to safety] [8].
9. European Medicines Agency. A Guideline on Summary of Product Characteristics (SmPC), Revision 2 (2009) [21].
10. CIOMS VI – Management of Safety Information from Clinical Trials (2005) [1].
11. FDA Reviewer Guidance. Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review (2005) [22].
12. FDA Guidance for Industry: Premarketing Risk Assessment (2005) [23].
13. International Conference on Harmonization (ICH) M4E: Common Technical Document for the Registration of Pharmaceuticals for Human Use - Efficacy (2002) [24].
14. ICH E9: Statistical Principles for Clinical Trials (1998) [25].

15. ICH E3 Guideline for Industry: Structure and Content of Clinical Study Reports (1995) [26].

Additional references used to inform recommendations regarding AEs in general have been cited throughout the document.

9. Considerations

Members of the Analysis and Display White Paper Project Team reviewed regulatory guidance and shared ideas and lessons learned from their experience. Draft white papers were developed and posted in the PHUSE wiki environment for public comments.

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

10. Recommendations

10.1. General Recommendations

This section contains general recommendations for analyses and displays that apply to safety in general, not specifically to AEs, but applies to AEs as well.

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 [11]. This white paper does not attempt to resolve this debate. As noted in the FDA Clinical Review Template (eg, Section 7.4.2 of Attachment B) [8], 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. In the example tables (eg, Table 10) in the FDA Clinical Review Template [8], there is a note that while p-values are not used for hypothesis testing in safety, it is useful to identify events with p<0.05 for drug/placebo comparisons. International Conference on Harmonization E9 Section 6.4 [25] notes that descriptive statistics are generally used for safety, with confidence intervals wherever it aids in interpretation. International Conference on Harmonization E9 [25] also mentions that p-values are sometimes useful as a flagging mechanism to highlight differences worth further attention. Ma and colleagues [14] states 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 that these are not helpful as a tool for reviewing the data, they can be excluded from the display. 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 particular threshold is met (eg, p<0.05). 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 emphasized in a February 2016 statement from the American Statistical Association [27]. Although certain statistical methods are recommended in this white paper for confidence intervals (for teams that choose to include them) alternate methods can be considered.

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 over-interpretation 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 over-interpreted. As noted in ICH E9 Section 6.4 [25], 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, and the February 2016 statement from the American Statistical Association [27] provides helpful guidance to avoid over-interpretation.

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; medical judgment always needs to be applied [28].

When only percentages are included in displays (eg, 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 in their head. Some physicians may place an emphasis on the most frequent events, while others may place an emphasis on those that have the largest risk ratios (eg, the treatment percent is at least 2 times the placebo percent). Without p-values or confidence intervals, strength of evidence for a drug/placebo difference would need to be inferred. Other physicians may only focus on those events that they believed prior to review to have biological plausibility.

Statistical Methods and Multiplicity

If a sponsor or compound team decides to include p-values as part of their practice to facilitate focus for signal detection, the next decisions are what method and if and how to adjust for multiplicity. Typically, for individual studies, the Fisher's exact test is used. For integrated summaries, the Cochran-Mantel-Haenszel test for general association adjusted for study is a reasonable option, but other choices may make sense especially for rare AESIs [29]. As noted by Wisniewski and colleagues [30], while ICH E9 [25] recommends statistical adjustments for multiplicity when applying hypothesis tests to a large number of safety variables in clinical trial data, multiplicity adjustment has also been described as counterproductive [8] and are often avoided. Crowe and colleagues [11] mentioned two possible

methods to tackle multiplicity – the Double False Discovery Rate multiple adjustment procedure [31] or a Bayes method suggested by Berry and Berry [32]. The Bayes method utilizes a Bayesian hierarchical model that can use the MedDRA hierarchy, thereby borrowing strength from similar terms. It has been suggested that it performs well [33]. Since this version of this white paper does not take a position on whether to even include p-values, we do not take a position on any multiplicity method. For now, since we believe p-values are primarily used for knowing which terms have similar strength of evidence (not hypothesis-testing), we think avoiding multiplicity methods is acceptable. However, we are intrigued by methods that can facilitate the identification of terms most worthy for the medical reviewer to consider.

Choice of Comparative Metric

When summarizing AE data between treatments for signal detection, there are several choices for a comparative metric (when included). The most common choices (for clinical trials) are the risk difference (the difference between two proportions), risk ratio (the ratio of one proportion to another), and odds ratio (the ratio of one odds divided by another, where the odds of an event is a proportion divided by one minus the proportion). The risk difference is an absolute measure, while the risk ratio and odds ratio are relative measures. There are advantages and disadvantages across these choices [4, 14, 34]. As noted by Zhou and colleagues [4], the risk difference can more directly reflect the magnitude of patients that could be affected by a risk. However, relative metrics are sometimes used as a flagging mechanism for identifying events that may warrant further investigation (eg, identifying events occurring at least two times more frequently with drug versus placebo). We chose to display risk differences in our example tables. Replacing the risk difference with an alternate metric (eg, risk ratio or odds ratio) would also be acceptable. For AESIs, providing multiple metrics (eg, one absolute metric and one relative metric) is often warranted. However, for general safety signal detection providing one metric is generally sufficient.

Importance of Visual Displays

Communicating information effectively and efficiently is crucial to detecting safety signals in a timely manner and enabling 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 detecting key safety signals and improving the ability to make clinical decisions [35, 36]. They can also facilitate identification of unexpected values.

Standardized presentation of visual information is encouraged. The FDA/Industry/Academia Safety Graphics Working Group was initiated in 2008. More information on this group can be found at *https://www.ctspedia.org/do/view/CTSpedia/StatGraphHome*. This working group was formed to develop a wiki and to improve safety graphics best practice. It has recommendations on the effective use of graphics for three key safety areas: AEs, electrocardiograms (ECGs), and laboratory analytes. The working group focused on static graphs, and their recommendations were considered while developing this white paper. In addition, there has also been advancement in interactive visual capabilities. While this white paper focuses on static displays, we do include some notes for areas where interactive visual capabilities would be beneficial and include snapshots of examples in Section 18, Appendix B.

Conservativeness

The focus of this white paper is clinical trials for which there are comparator data (either placebo or active comparator). One might be tempted to consider that it is always "conservative" to choose a method that leads to a higher number of patients being included in an analysis. This could be achieved in various ways, such as including a long post-drug follow-up time, using a low boundary as the upper reference limit for a laboratory value, or using a broad search for AE terms to ensure that all possible relevant terms are included in the definition of the safety outcome. However tempting these approaches may be, they can sometimes make it more difficult to identify safety signals in a comparison of drug to a comparator [8, 37, 38]. This is because overly broad inclusion of events (or patients with an event) can lead to an underestimation of the true relative risk because more events may be included in each arm that are unrelated to the true, but possibly unknown, mechanism of action, thus adding spurious "events" to each arm in the analysis.

In other words, it is not necessarily "conservative" to add more data, if the resulting estimate introduces bias and/or spurious findings that could dilute a signal. The intent of the proposed methodology is to identify meaningful safety signals for a treatment relative to a comparator group.

Number of Therapy Groups

The example TFLs show two treatment arms (eg, two dose arms) versus a comparator in this version of the white paper. Most TFLs can be easily adapted to include additional treatment arms or a single arm.

Multi-phase Clinical Trials

The example TFLs for individual studies show two treatment arms and a comparator within a controlled phase of a study. The example TFLs for integrated summaries show one treatment arm (assumes all of the treated arms are pooled) and a comparator arm within the controlled phase of the studies. Example TFLs for additional phases are considered out-of-scope in this version of the white paper. Many of the TFLs recommended in this white paper can be adapted to display data from additional phases.

Integrated Analyses

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 [22] 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 (eg, 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 (eg, including study as a stratification variable) is usually important. When the treatment-placebo randomization 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 (eg, Simpson's paradox; [13]). Creating visual displays or tables in which comparisons are confounded within study is discouraged.

One method that has been proposed by Chuang-Stein and Beltangady [12] is to create "adjusted percentages". Even when appropriate stratified methods are used for the summary metric, confidence interval, and/or p-value, summaries generally include the crude percentage, which can be confusing if a paradox exists because of unequal randomization. Methods have been suggested but are not yet part of normal practice. We believe that it is time to use adjusted percentages more routinely. While there are multiple methods for creating adjusted percentages, we encourage the inclusion of the study-size adjusted percentages in displays [13]. When the unadjusted percentages are impacted by such paradoxes, the adjusted percentages are likely to be more appropriate for summaries such as labeling.

Table 10.1 illustrates the phenomenon. It can be seen that in each study of Table 10.1 the percentages of patients with the AE of interest were the same. However, the crude pooled percentage shows that the 24% of patients treated with the new treatment experienced the AE, while 29% of patients treated with placebo experienced the AE. The study-size-adjusted percentages accurately reflect the lack of difference between the new treatment and placebo-treated patients. This paradox is not limited to any particular direction. It could cause signals to be missed or it could cause undue alarm.

Table 10.1.
Observed Incidence Proportions of Patients with a Particular Adverse Event in the New Treatment and Placebo Group in Three Trials Data Extracted from Table 1 of Crowe and colleagues [13]

	New Treatment, n/N (%)	Placebo, n/N (%)	Total patients in the study
Phase 2 Study	30/300 (10.0%)	10/100 (10.0%)	400
Phase 3 Study	133/700 (19.0%)	67/350 (19.1%)	1050
Phase 3 Study in Refractory Patients	200/500 (40.0%)	200/500 (40.0%)	1000
Percentages from Crude Pooling	363/1500 (24.0%)	277/950 (29.0%)	2450
Study-Size Adjusted Percentages	$\frac{400}{2450} \times \frac{30}{300} + \frac{1050}{2450} \times \frac{133}{700} + \frac{1000}{2450} \times \frac{200}{500}$ = 26.1%	$\frac{400}{2450} \times \frac{10}{100} + \frac{1050}{2450} \times \frac{67}{350} + \frac{1000}{2450} \times \frac{200}{500}$ = 26.2%	

N = the total number of patients in the group; n = the number in the group that experienced the event.

In addition to avoiding potential misinterpretations of pooled data, understanding whether the overall representation accurately reflects the review across individual clinical trial results is also important. It is likely common practice to review results by study for events that have percentages that appear different between the treatment group and placebo in an integrated summary. However, it might not be as common to review results by study for events that have percentages that appear similar between the treatment group and placebo in an integrated summary. In order to not miss potential safety signals, it is important to include methods that would identify a potential signal that might only be apparent in one of the studies (see the phototoxicity example in Section 7.1.3 in Attachment B of the FDA Clinical Review Template [8]). In order to detect such situations, tests for homogeneity could be included in the tables for integrated summaries (not shown in our example tables), interactive tools can be used to systematically review by-study results, and/or a careful review for potential signals in individual CSRs can be included as part of the overall safety review process.

10.2. Discussion of Adverse Event Definitions

As discussed in the FDA Reviewer Guidance [22], an adverse event is any untoward medical occurrence associated with the use of a drug in humans, whether or not it is considered study-drug related. An adverse drug reaction is an undesirable effect, reasonably likely to be caused by a study drug and it may occur as part of the pharmacological action of the study drug or may be unpredictable in its occurrence. Adverse drug reactions do not include all AEs observed during the use of a study drug, only those where there is some reason to believe a causal relationship exists between the study drug and the occurrence of the AE. Determining whether an AE is an ADR involves many factors (eg, biologic plausibility, clinical impressiveness of any individual case, the statistical assessment of the strength and magnitude of the observed effect, the observed dose relationship, the severity of the event, the consistency of findings across studies, the consistency of findings from similar events, consistency of findings from similar compounds and clinical relevance) [10, 14, 39].

Serious adverse events are AEs occurring at any dose, whether or not they are considered study-drug related, that result in any of the following outcomes:

- Death
- A life-threatening AE
- Inpatient hospitalization or prolongation of existing hospitalization
- A persistent or significant disability/incapacity
- A congenital anomaly or birth defect
- Important medical events that may not result in death, be life-threatening, or require hospitalization may be considered SAE when, based upon appropriate medical judgment, they may jeopardize the patient and may require medial or surgical intervention to prevent one of the outcomes listed above

Non-serious adverse events are all AEs that do not meet the above criteria for “serious”.

The precise definition of a **treatment-emergent adverse event** varies across documents and all lack necessary detail for consistent implementation across the industry. Section 7.4.1 of the FDA Clinical Review Template [8] notes that treatment-emergent signs and symptoms are those not present at baseline or not present at the severity seen on treatment. Several other authors have discussed more detailed definitions for treatment emergence [10, 40]. As noted in these references, variation in collection methods for AEs introduces additional variation in the eventual interpretation of treatment emergence as well. A future white paper is planned to illustrate more thoroughly how variations in collection and definitions potentially affects the summary of TEAEs. Recommending a specific collection method and more detailed definition of treatment emergence is considered out-of-scope for this version of the white paper but is very much in scope for the overall computational science collaboration. It is assumed TEAEs are identified in analysis datasets and available for summaries/analyses. The specific detailed definition should be documented in the compound’s PSAP [11, 41-43], or similar planning document that is developed early in a compound’s lifecycle and shared with regulatory agencies for feedback as needed. The specific method should also be included in SAP(s) and protocol(s) methods sections, as appropriate.

10.3. Adverse Event Data Collection

Adverse event data can be collected for a study/product in several ways. Some methods for obtaining AE data include open-ended questioning, specific solicitation of particular AEs, and checklists. Especially for open-ended questioning, the instructions for when to include an event for collection become important for truly understanding “adverse events” and “treatment-emergent adverse events”. The method of obtaining information should be considered carefully as there may be limitations in interpretation depending on the collection approach. Across studies, consideration should be made to proactively collect AE data consistently at least through specific periods of a compound’s lifecycle. As noted in FDA’s guidance on safety data collection [16], variation in collection is encouraged between early-phase studies versus late-phase (eg, postmarketing) studies.

As noted in Section 10.2, recommending a specific method for collection and/or specific instructions associated with collection is considered out-of-scope for this white paper but is in scope for the overall FDA/PHUSE computational science collaboration. A newly formed project team named “Best Practices for Data Collection Instructions” has just started tackling some of the challenges in data collection consistency.

While it is out-of-scope for this white paper to provide specific recommendations, below is a brief summary of some areas where variation occurs (with some thoughts included) to highlight the issue:

- Events that start after informed consent obtained but before treatment initiation – For most AE collection methods, events that start after informed consent date but before treatment initiation are captured in AE collection, but some collection methods may start at the time a treatment phase begins. In addition, some collection methods may collect all AEs after informed consent but before treatment initiation, while others may limit the AEs during the time window to include only events deemed to be causally associated with or due to protocol-mandated procedures.

- Events that start before treatment initiation and worsen in severity after treatment initiation are captured in routine AE collection; however, the details vary. See Section 10.5 for further discussion of event severity.
- Events that start before treatment initiation but improve after treatment initiation – Improvements are captured in some collection methods but not in many other collection methods. Having a collection method that captures improvements (ie, all changes in severity) does provide better data for individual case reviews (eg, enables the graphical summaries of AEs over time as shown in Section 18, Appendix B). Having a collection method that captures changes in severity over time also provides more complete data for potential ad-hoc exploration that may not be possible otherwise (eg, potential relationship between agitation/activation and suicidal thoughts and behaviors). See Section 10.5 for further discussion of event severity.
- Event relatedness – In most collection methods, it appears event relatedness is collected. See Section 10.6 for further discussion of event relatedness.
- Serious event timing – For some collection methods, the only thing captured is if an event was serious at some point during the study. In other collection methods, more details are known about serious events. Of note, for multiphase studies, it is likely to be important to know when an event becomes serious to enable summaries of SAEs by phase.
- Definition of serious – In some collection methods, serious is strictly determined by the definitions per regulatory guidance (eg, *http://www.fda.gov/Safety/MedWatch/HowToReport/ucm053087.htm*). In other collection methods, some sponsors add additional criteria to what should be considered serious. Sometimes, these additional criteria are added simply to provide a means to receive more detail on individual case descriptions. In these cases, we recommend specifically designed follow-up forms for AESIs instead of merely defining these as "serious".
- Serious events occurring after study completion – In some collection methods, serious events occurring after study completion are included in the study clinical trial database, while in other collection methods they are included in alternate data systems. We recommend that the clinical trial database include all serious events that occur during a study and not include serious events that occur after the study is considered complete for the patient. Predefined follow-up phases should be considered part of the study and part of the design of the study.
- Intermittent events (eg, headaches) – These are sometimes captured as one event over a long period, or sometimes each event is captured. These variations make summaries of the number of events difficult to interpret.
- Events that are part of the disease – Prespecified events considered part of the disease are captured in the Clinical Events CDISC domain. However, events that might be related to the disease that were not prespecified may be captured inconsistently.

While we will defer to the newly formed project team for more robust recommendations, we will make the following assumptions for purposes of recommending displays: all events starting at informed consent date will be captured (with programming code that includes treatment start date to be used to determine treatment emergence); event severity will be collected either as mild, moderate, severe or using Common Terminology Criteria for Adverse Events (CTCAE) grade severities; and severities over time will be captured (including improvements).

10.4. Adverse Event Categories and Preferred Terms

Investigator-reported AEs usually include verbatim terms. One commonly accepted approach to placing verbatim terms into appropriate AE categories prior to analyses includes mapping to a standard dictionary of PTs, one approach being MedDRA (*http://www.meddra.org*). The Medical Dictionary for Regulatory Activities is widely used across the industry in nononcology studies. In oncology studies, CTCAE grades are commonly used, which includes MedDRA mapping and definitions for severity grades (see Section 10.5). The TFL analyses proposed in this document assume that investigator-reported AEs are mapped appropriately to AE categories.

As noted in Attachment B, Section 7.1.2 of the FDA Clinical Review Template [8] and the EU SmPC Guideline [21], categorization that is too granular can result in underestimation (eg, somnolence, drowsiness, sedation, and sleepiness probably all refer to the same event), and too broad of categorization can result in the dilution of more significant events (eg, loss of consciousness and syncope subsumed under hypotensive events or hypotension). The Medical Dictionary for Regulatory Activities provides several levels of granularity – lowest level term (LLT), PT, high-level term (HLT), high-level group term (HLGT), and system organ class (SOC). As noted in Section 6.2 in ICH E9 [25] and Section IV (f) in the CIOMS VI report [1], the "preferred term" is generally recommended for reporting purposes. Additionally, using the MedDRA PT for reporting is currently predominant industry practice. For now, we support the use of MedDRA PT (nested within SOC) in the standard set of static displays but believe additional steps should be taken to address potential over-granularity, and we believe further research is needed in this area. We currently believe an additional static display incorporating MedDRA PT nested within HLT and SOC could facilitate an assessment that is less granular than the MedDRA PT and is recommended for integrated summaries. However, since the MedDRA HLT level does not address all potential over-granularity issues (eg, HLTs do not combine PTs that are in different SOCs), a review of AEs for signal detection should include a review of whether excessive splitting exists in the data that would potentially require additional grouping of terms. Such a review could be done manually (by reviewing the entire TEAE table reported using PTs), and/or by utilizing interactive tools that allow for selecting terms to group interactively. The FDA uses an interactive display tool called JReview as well as an internally developed tool called MedDRA-Based Adverse Event Diagnostics (MAED) that allows reviewers to create custom queries with specific PTs to look for safety signals. At the time of submission (when decisions are made with respect to determining ADRs to communicate in labeling), a discussion in the Summary of Clinical Safety on how events were reviewed to address these potential pitfalls and the result of the review is often warranted.

Alternatively or additionally, sponsors can create and maintain custom groupings. At least one company [10, 33, 44] maintains company-specific "MedDRA Labeling Groupings" to address potential over-granularity for reporting frequencies in labeling to use across compounds within their company. Unlike HLTs, these groupings can be made to combine terms that are essentially the same event but not necessarily in the same SOC (eg, Mallick

[44] used the following PTs to represent the clinical concept of anaemia: Anaemia, Erythropenia, Haematocrit decreased, Haemoglobin decreased, Packed red blood cell transfusion and Red blood cell count decreased). These custom groupings are potentially useful for improving signal detection. According to some research on the topic, it appears there is some improvement in signal detection with the use of such groupings [30, 33]. If a-priori systematic grouping of similar terms (that can span across SOCs) is deemed valuable for signal detection, perhaps the PHUSE Working Group collaboration could facilitate making use of such groupings operationally available for cross-industry use (eg, networking with the MedDRA Maintenance and Support Services Organization (MSSO) and/or having a cross-industry group support the effort and store the result in a public area such as the PHUSE Repository). We consider improvement in this area a high priority in the overall standardization effort. Of note, MedDRA provides an additional mechanism in which to group terms –SMQs. Standardized MedDRA queries have been developed for many clinical concepts. Generally, SMQs include a relatively large list of PTs related to a particular medical concept (which can span across different SOCs). Many (though not all) medical concepts are covered and they continue to evolve. They tend to be very helpful when searching for events that potentially point to a common pathophysiology but often require some level of review for identifying true cases of interest. The SMQs help somewhat to address the potential over-granularity of MedDRA PTs, but these groupings are generally broader than the attempts to group PTs that are essentially the same event. Generally, their purpose is to assist in the assessment of AESIs. If an SMQ exists for a particular AESI, it is recommended to use the SMQ since there are several benefits to using them rather than developing a custom definition. The main advantages 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 recognized by regulatory agencies. Because newer MedDRA versions have more SMQs, studies that used old versions of MedDRA should be upconverted to a newer version of MedDRA for a more comprehensive integrated safety review.

10.5. Adverse Event Severity

As noted in the CDISC SDTM Implementation Guide, severity seems to be collected in most cases. Severity is generally available for an event when it is first reported and/or when it is most severe. However, the reporting of severities over time (eg, all changes including improvements) is inconsistent. As noted in Section 10.3, we assume severities over time are available. If they are not available, then some of the displays in this white paper will not be possible or will have to be adjusted. Of note, in many cases, severity is incorporated into programming logic for determining treatment emergence. Therefore, before a study team decides not to collect severity, a thorough understanding of potential downstream implications should be understood. It appears that for most nononcology studies, severity is collected subjectively as mild, moderate, or severe, and for most oncology studies, severity is collected more objectively using CTCAE grade definitions with 4 or 5 levels for some terms. Standardized CTCAE severity levels were developed for oncology studies since oncology patients experience many AEs and such severity granularity is considered useful. Since the development of CTCAE severity levels, some have suggested that it should be used beyond oncology, with the idea that having more severity levels with objective criteria would be more useful than three subjective levels. Alternatively, others have suggested that collection of severity level is not useful with the idea that knowing seriousness and knowing whether the event led to treatment discontinuation are sufficient indicators of severity. In general, we do not recommend the added burden to use CTCAE severity definitions in nononcology indications, except perhaps for specific AESIs, if applicable. In this version of the white paper, we provide mock shells for both – mild/moderate/severe (generally used for nononcology) and CTCAE-defined severity levels (generally used for oncology). However, for nononcology studies, a case can be made not to create a static table summarizing severity levels. It might be better to use the data for individual case reviews and/or provide the ability to review severity levels into interactive display tools for exploration purposes instead of creating static tables. For now, we will recommend a static display but this is certainly one area for further thought.

10.6. Adverse Event Relatedness Assessment by the Investigator

As noted in the FDA Guidance on Safety Reporting Requirements [9], the investigator's assessment of causality is not required for nonserious AEs by the regulations, although many sponsors require it in protocols. As discussed in the FDA Clinical Review Template [8], the relatedness assessment made by the investigator generally is not considered useful. In the CIOMS VI report [1], there is a recommendation only to collect relatedness for serious events. However, in the Japan guidance document [20], there is a mock table that includes columns providing percentages of the AEs considered related (side-by-side with all-causality AEs). This table with both all-causality AEs and treatment-related AEs are generally used in labeling for Japan. While we concur with the CIOMS VI working group that collecting relatedness only for serious events makes sense, we recommend that the relatedness assessment made by the investigator be collected for all AEs to meet the current expectations of the Pharmaceuticals and Medical Devices Agency (PMDA). We do not believe summary tables for AEs considered related by the investigator are necessary for individual studies, nor necessary in submission documents to regulatory agencies other than the PMDA. In addition to the expected tables for the PMDA, the data can be used for individual case reviews and/or provided in interactive display tools for exploration purposes.

Of note, the specific levels of relatedness are not defined in CDISC/CDASH. Thus, when collected, collection varies (eg, yes/no, yes/no/unknown, no/probable/possibly/likely).

The relatedness assessment by the investigator is different from the United States requirements for expeditiously reporting SAEs to the FDA. Per the FDA Final Rule (21 CFR 312.32) and guidances on Safety Reporting Requirements [9, 45], sponsors are required to expeditiously report any SAEs that are unexpected and for which there is evidence to suggest a causal relationship between the study drug and the SAE (ie, the SAE is suspected to be an ADR). Under the Final Rule, even if an investigator identifies an event as study-drug related, but the sponsor finds no evidence of causality, the sponsor should not report the event on an expedited basis to the FDA [46]. The idea is that a sponsor would have more information than an individual investigator would have (eg, SAEs reported across study sites and across multiple studies, the study drug's

mechanism of action, class effects, aggregate analyses). While the investigator's view is important for the sponsor to consider when reviewing SAEs, all information is used for determining potential causality.

10.7. Adverse Events after Stopping Study Drug

Adverse events that occur after stopping study drug (eg, treatment under study and comparators) are collected for many reasons. In many cases, safety follow-up phases are included to monitor patients for a period after study drug is stopped. Follow-up phases are often included to determine if there are safety concerns when patients stop study drug, to allow for event collection through the time the study drug is expected to be out of the patients' systems, and/or to monitor certain patients that have a certain event until their condition improves. In addition, study designs that keep patients in a study (for the entire planned length of time) after stopping study drug are becoming more popular as part of missing data initiatives [47, 48]. In these cases, patients can be off study drug for an extended period. While these missing data initiatives pertain primarily to efficacy, these study designs now create confusion as to how to report events occurring during double-blind treatment but after study drug was stopped.

Some guidance documents contain advice on how long to collect AEs post last dose of study drug (eg, one dosing interval post last dose, 30 days post last dose, or a certain number of half-lives after the last dose). It is extremely important to document during collection the best estimate of the last date study drug was taken, as well as the dates when AEs start/end/change severity, so that an accurate determination can be made of timing of events relative to the treatment dates.

Regardless of how long AEs are collected post last dose of study drug, there is currently no standard approach to their analysis. Different methods for handling TEAEs occurring after study drug is stopped are likely one of the biggest causes of variation across the industry, and what medical reviewers request of sponsors seems to vary. Some researchers have the perception that it is standard practice to include TEAEs that occur up to a predefined period of time after study drug is stopped (eg, within 2 weeks of last dose of drug, within 30 days of last dose of drug, or within a certain number of half-lives after the last dose). If a sponsor or review division has a business rule, it would need to be applied with caution, as the rule may not be appropriate in all cases. We believe getting alignment on some basic principles on this topic is a high priority for future work in the overall standardization effort.

Our recommendations for how to handle TEAEs occurring post study drug depend on the design of the study and the expected characteristics of the study drug(s). The following provides some general principles:

- All TEAEs that occur during a study should be included as part of a summary in some manner.
- When summarizing TEAEs for the controlled-treatment phase (eg, study drug versus placebo), if TEAEs that occur after stopping study drug up to a certain number of days (eg, within 2 weeks of last dose of drug) are included, then TEAEs should be included for all treatment arms using the same number of days. For example, bias would be introduced if TEAEs after stopping study drug are included for some arms (eg, treatment arm) but not for other arms (eg, placebo arm).
- When summarizing TEAEs for the controlled-treatment phase (eg, study drug versus placebo), TEAEs occurring after the treatment phase should not be included if patients could switch to a different active treatment. For example, if the study design were a double-blind phase followed by an open-label phase, placebo patients would switch to drug after entering the open-label phase. In this design, adding a set number of days after study drug end date or treatment phase end date (eg, within 2 weeks) could result in treatment-related events being attributed to placebo. A rule could be added that TEAEs after a patient switches treatment would be excluded but that could introduce bias (as discussed in the previous bullet) if that essentially means days are added for the treatment group and not the placebo group. For designs that have an open-label phase immediately after a double-blind treatment phase, we recommend not including TEAEs after the treatment phase (even if the TEAE occurred shortly after last dose date from the treatment phase) for summarizing TEAEs for the treatment phase. The TEAEs would not be ignored, just part of a separate display.
- When reporting TEAEs adjusted for time (eg, exposure-adjusted incidence rate, exposure-adjusted event rates), consistency is required between the numerator and denominator. For example, if TEAEs during a follow-up period were planned to be included in the analysis of the treatment period, then the calculation of time would need to include the follow-up period. Note that if TEAEs after stopping study drug are included, the EAIR may be lowered if the true rate of the event decreases after stopping study drug.
- For compounds and designs in which there is an expectation for new TEAEs to occur after stopping study drug (eg, return of disease symptoms; introduction of a concomitant medication, if allowed per protocol; and/or anticipated discontinuation or withdrawal effects of the study drug), separate TFLs should be created for emergent events occurring "during treatment" (TEAEs) and "off treatment" (sometimes called follow-up emergent events). This enables the researcher to distinguish between drug-related safety signals and safety signals that are more likely related to discontinuation of drug. When summarizing TEAEs for a treatment phase in this case, including TEAEs after stopping study drug would not be recommended.
- In study designs that keep patients in a study for the entire planned length of time (following the schedule of events as if on study drug) even after stopping study drug, additional clarity is needed for how events after stopping study drug would be handled. When summarizing TEAEs for a treatment phase, the inclusion of TEAEs occurring after stopping study drug but still occurring during the treatment period would generally not be recommended. For compounds with very few patients stopping study drug and little concern for return of disease symptoms and/or discontinuation/withdrawal effects, including all TEAEs while in the treatment phase but not "on treatment" would be acceptable.
- Generally, at least some TFLs that include data from follow-up phases and/or time off study drug will be required, but not usually as many as those done for during treatment and not necessarily in the same format as provided in this white paper. For some compounds (eg, compounds with a long half-life compared to the duration of the study; compounds used for a short time, such as antibiotics), a more complete set of TFLs including such data may be required. The ease of interpretation from such TFLs will vary depending on the compound,

disease, and study design aspects, such as the half-life of the compound, the likelihood of taking alternate therapy, or the allowed concomitant drugs during the observation period.
- Except for AESIs, conducting sensitivity analyses for various ways to handle TEAEs that occur after stopping study drug is generally unnecessary. For a submission, the possible exception is that it may make sense to include both an integrated summary of all TEAEs while on study drug (any patient ever to take the investigational study drug during any phase of any study, not including TEAEs occurring after stopping study drug) AND a summary of all TEAEs while on investigational study drug that includes TEAEs occurring after stopping study drug including those occurring in follow-up phases. This 2-column table may be an easy-to-implement way to assess the different approaches.
- While not necessary, it is usually best that the methods for considering TEAEs after stopping study drug for TEAE summaries of the treatment period be consistent with methods for SAE summaries of the treatment period. However, how to handle TEAEs occurring after stopping study drug may vary for AESIs (especially for rare, more serious AESIs that could have late-onset).
- While not necessary, it is usually best that methods for handling TEAEs after stopping study drug for TEAE summaries be conceptually consistent with methods considered for analyses of laboratory, vital signs, and ECGs.

See a few examples of how these principles would be applied:

Example 1 (typical example from various therapeutic areas): Consider a study that has a double-blind treatment phase (study drug versus placebo), then a no-drug follow-up phase based on the half-life of the study drug. The follow-up phase is relatively short compared to the treatment phase. Patients who permanently discontinue study drug are encouraged to remain in the study following the normal schedule of events as if they remained on study drug (so these patients can be off study drug for a very long time). There is concern for a return of symptoms and/or withdrawal effects after stopping the study drug. In this case, we recommend separate summaries for the treatment and no-drug follow-up phases. Additionally, for those patients who permanently discontinue study drug but remain in the study following the normal schedule of events, TEAEs occurring after permanently discontinuing study drug should be included in the summary of follow-up events (not part of the treatment phase summary). The summary for the treatment phase would include study drug versus placebo. The summary for the no-drug follow-up phase could still be separated by study drug and placebo (based on what the patient was on during the treatment phase) even though all patients are off study drug.

Example 2 (typical example from diabetes and autoimmune studies): Consider a study that has a double-blind treatment phase, then an open-label treatment phase, then a no-drug follow-up phase. Patients who permanently discontinue study drug are encouraged to remain in the study following the normal schedule of events as if they remained on study drug. There is concern for a return of symptoms and/or withdrawal effects after stopping study drug. In this case, we recommend the summary of TEAEs for the treatment phase (study drug versus placebo) for the CSR to include only the TEAEs during the treatment phase. If TEAEs that occur within a specific number of days after last dose of study drug were included, TEAEs that occur during the open-label treatment phase might be included and would create bias. Additionally, for those patients who permanently discontinue study drug during the treatment phase but remain in the study following the normal schedule of events, AEs that occur after permanently discontinuing study drug should not be included. A separate summary table (for study periods combined) can be created with 2 columns: 1) including TEAEs while on study drug (at any time during double-blind or open-label) and 2) including TEAEs while on study drug or after study drug but still during the study (at any time during double-blind, open-label, or no-drug follow-up). For the submission (integrated with other studies), a table addressing a potential withdrawal effect could include all TEAEs first occurring or worsening after the treatment end date but within a specified number of days (based on the half-life of the drug).

Example 3 (typical example from oncology studies): Consider a study that has a double-blind treatment phase (study drug versus placebo), then a no-drug follow-up phase based on the half-life of the study drug. The follow-up phase is relatively short compared to the treatment phase. Patients who permanently discontinue study drug enter the follow-up phase. There is little concern for a return of symptoms and/or withdrawal effects of the study drug during the follow-up phase. In this very specific case, we believe it is acceptable that the summary of TEAE percentages (study drug versus placebo) for the CSR include the TEAEs occurring during the treatment phase and follow-up phase (combined). For the submission (integrated with other studies), at least one display will likely be required to address the potential for a withdrawal effect (per ICH guidance) – a summary of events first occurring or worsening after the treatment end date. Of note, if the phases are combined, the interpretation of the results should be clear about the periods the estimates are based. For example, if the treatment period is 12 months and the follow-up period is 1 month, estimates correspond to 12 months of treatment plus one month of no drug, not 13 months of treatment.

10.8. Calculating Percentages using Population-Specific Denominators

As noted in the FDA Clinical Review Template (page B-76, Table 10, Footnote 7) [8], percentages for sex-specific AEs (eg, impotence) should be determined using the appropriate number of sex -specific patients as the denominator, and it should be indicated with a footnote. Recommended footnotes for this purpose are provided in the table shells in this white paper. The MedDRA MSSO has been providing a list of sex-specific AEs for each MedDRA version and the list can be used for determining which events should have a sex-specific denominator (although the list was created with a different purpose in mind (to improve data quality) (see *http://www.meddra.org/sites/default/files/guidance/file/intguide_19_0_english.pdf*). However, the MedDRA MSSO has decided to discontinue maintenance of these lists (see *http://www.meddra.org/paediatric-and-gender-adverse-event-term-lists*), so lists will no longer be available as of version 19.1. Thus, companies may need to maintain their own lists, determine which events require sex-specific denominators manually, or find another process. Alternatively, by-gender tables can be created. By-gender tables have the advantage of being easy to implement and seeing by-gender analyses might be of interest anyway. However, having by-gender tables for TEAEs, AEs leading to discontinuation of study drug, and SAEs have the disadvantage of increasing substantial volume in a CSR and the general feeling is to save a lot of subgroup exploration

for the integrated stage (when more data are available for subgroup analyses). In addition, having sex-specific denominator adjustment within the core table allows for sorting by frequency to be done more appropriately (eg, without sex-specific denominator adjustment, the sex-specific terms fall far lower in the sort). Our PHUSE project team plans to continue discussion on this topic to help find ways to make such sex-specific denominator adjustment easier to implement. For example, a list can be created and maintained by a cross-industry group and either provided to MedDRA or stored in the PHUSE Script Repository.

Other populations may warrant population-specific denominators. For studies with a mix of pediatric and adult patients, pediatric-specific denominators should be considered. The MedDRA MSSO has been providing pediatric-specific lists in addition to gender-specific lists, but they have been discontinued as of version 19.1. Theoretically, percentages for many events can be created using even more specific denominators from only those demographics that can have the event (eg, menstrual cramps would only apply to premenopausal women). In practice, unless the event is of particular interest, we recommend only attempting such adjustments for relatively broad demographic groups (eg, male, female, pediatric populations).

10.9. Alternate Methods when Percentages are Biased

When comparing TEAEs between treatment groups in a fixed duration controlled treatment period (ie, protocol plans for same follow-up duration in each arm), percentages are generally sufficient for safety signal detection. For combined analysis of multiple studies, study-size-adjusted percentages are a natural analogue [12, 13]. However, there are some cases in which comparing percentages may over or underestimate the treatment effect.

One circumstance in which percentages are problematic is when the observed time is different among the treatment groups within a study (or within any of the individual studies in a combined analysis of multiple studies). This could be because of study design features (such as including 'rescue' to another therapy) or differential dropout among the groups. Longer-term controlled treatment periods will have greater chances of uneven dropout among treatment groups than shorter-term trials, but in cases where dropout distribution and rates are similar and not overly high, percentages are sufficient and additional metrics are generally unwarranted. A percent is valid even for events with nonconstant hazard. Events with nonconstant hazard have varying risks for developing them over time. An example would be the risk for severe allergic reactions to medicines, which tends to be higher earlier versus later in the treatment period.

Percentages may be problematic even if the study design is such that patients are intended to be followed for the same duration and dropout is similar in both groups. In this case, observed percentages may underestimate what would have been seen if there had been no dropout. Generally, for signal detection, we are interested in the difference (absolute or relative) between groups, and even with similar dropout, the difference could be somewhat impacted (though not to the same extent as with dropout that is different among the groups).

In situations where percentages are deemed problematic, other metrics, such as exposure-adjusted incidence rates or time-to-event, may be appropriate. Of course, each of these can give biased results in some situations, so careful thought should be given to the appropriate metric. While summarizing multiple metrics for AESIs may be warranted, we do not recommend using multiple metrics for general safety signal detection, as it makes review of the data unnecessarily complex.

When the time at risk of an AE is substantially different among treatment groups (within individual studies), the exposure-adjusted incidence rate may be a useful metric. Exposure-adjusted incidence rate is defined as the number of patients with a TEAE divided by sum of person-time at risk. Note that there is a study-size adjusted version of it that can be used for combining exposure-adjusted incidence rates from multiple studies [13]. For patients who had the TEAE of interest, person-time at risk is the time to event (from start of therapy). For patients who did not experience the TEAE of interest, person-time at risk is their follow-up time during the interval. The total person-time at risk is the sum of all the patients' person-time at risk. The most common metric for comparing exposure-adjusted incidence rates between treatment groups is the exposure-adjusted incidence rate ratio (which is the exposure-adjusted incidence rate in one group divided by the exposure-adjusted incidence rate in the other). Other metrics, such as exposure-adjusted rate difference (the difference of two exposure-adjusted incidence rates) are also possible.

Exposure-adjusted incidence rate methods implicitly assumes that "hazard" (which can loosely be thought of as the rate of an AE per unit time) for the AE in question is constant across time, an assumption that isn't always appropriate. By using total person-time at risk in the denominator, it effectively assumes that a large number of patients with short-term follow-up contrinbute the same information as a smaller number of patients with longer-term follow-up. Hence, exposure-adjusted incidence rates will give misleading results for events in which the hazard is not constant across time. As an example of nonconstant hazard, the risk for severe allergic reactions to medicines is generally higher earlier versus later in the treatment period. See Kraemer [49] for a discussion of how exposure-adjusted incidence rate can be a misleading statistic.

For long-term studies, it may be helpful to create exposure-adjusted incidence rates and exposure-adjusted incidence rate ratio (IRR) summaries/analyses over discrete periods. This allows visualizing how the exposure-adjusted incidence rate and how the treatment effect (exposure-adjusted IRR) changes over time.

Cox proportional hazards methods may also be useful for signal detection, instead of exposure-adjusted IRR, when percentages are deemed inappropriate. The main metric for signal detection with these methods would be the hazard ratio. While Cox proportional hazard methods do not require a constant hazard in any of the treatment groups, they do assume that the hazard in each treatment group is proportional to that of any other in the same study. This means that they may also be inappropriate for ADRs that tend to occur early or not at all, or for ADRs that only tend to occur after lengthy exposure to study drug.

Kaplan-Meier curves and/or hazard curves can help understand how the incidence changes over time and can be incorporated into interactive display tools. Note that while some of these

methods are appropriate for AESIs, it is generally unnecessary to subject the entire list of TEAEs to multiple methods for signal detection. Kaplan-Meier curves do not require any assumptions about the hazard being constant or not, nor about the treatment effects being proportional. They should generally be used for individual studies, not after pooling multiple studies' results, because of the possibility of misleading results (due to issues with pooling results from multiple studies with inconsistent randomization allocation as discussed in the "Integrated Analyses" subsection in Section 10.1). These methods tend not to be useful for extremely small numbers of cases.

Another instance in which percentages are problematic is when the event is so common that most patients will experience the event given a long enough observation period. For example, many drugs to treat diabetes have hypoglycemia as an AESI. With very long-term follow-up, nearly 100% of patients will have at least one episode of hypoglycemia. In this and some other circumstances, it is sometimes of interest to count the number of events as opposed to the number of patients with events (eg, number of hypoglycemic events in a diabetes study) and methods for handling recurrent events [50] might be more appropriate than metrics using number of patients. As with the exposure-adjusted incidence rate, this method (the exposure-adjusted event rate) assumes the risk for the AE is constant over time.

If recurring events on the same patient are of interest, special analyses may be warranted. See Ma and colleagues [14] and Bender and colleague [51] for additional discussion on methods for time of event onset, recurrent events, and methods when there is differential dropout among treatment groups.

11. Tables and Figures for Individual Studies

11.1. Recommended Displays

Throughout this section, recommended displays are provided. The displays are examples in which there are two treatment arms (low dose and high dose) and a placebo arm. The tables can be modified as needed under different scenarios. The examples also include 95% confidence intervals. As noted in Section 10.1, if a sponsor decides not to utilize confidence intervals as a tool for review, they can be deleted. In addition, if a sponsor decides to utilize p-values as a tool for review, they can be added.

As noted in ICH E3 [26], a brief summary of AEs is expected. To provide the basis for a brief summary, an overview table is very helpful (Table 11.1) and is recommended. It provides a high-level summary of the more detailed tables for AEs. For a more detailed summary of AEs, Table 11.2 through Table 11.6 and Figure 11.1 are recommended. Table 11.5 generally applies to non-oncology and Table 11.6 generally applies to oncology. Table 11.3 provides a summary of all TEAEs, and is sorted by SOC as recommended in ICH E3 [26] and in the example table in the FDA Reviewer Guidance [22]. As noted in the FDA Clinical Review Template [8] and ICH E3 [26], a summary of the most common TEAEs is expected. According to the FDA Clinical Review Template [22], common TEAEs are those TEAEs generally occurring at a rate of 1 percent or more; however, a higher cut off than 1 percent may be considered if doing so does not lose important information. Although a table is typically produced for the most common TEAEs (by simply subsetting on the table of all TEAEs; Table 11.4), Figure 11.1 is recommended as a more visual, user-friendly way to view the data. However, when the study is relatively small, making the confidences intervals quite wide, the visual becomes less user-friendly and a table is then recommended (Table 11.4). Table 11.5 or Table 11.6 is recommended to display maximum severity or maximum CTCAE grade (depending on collection). ICH E3 [26] notes that such a display may be useful for overall interpretation of AEs data. See Section 10.5 for further discussion regarding severity.
As noted in the FDA Reviewer Guidance [22] and ICH E3 [26], a listing of deaths is expected and is recommended (Listing 11.1). As shown in the example listing, we recommend sorting by treatment. However, for multi-phase studies where patients are treated with different treatments across the phases, the listing can be modified to be sorted by patient ID, with treatments as a column. A summary table of deaths (Table 11.2) would also be recommended for indications in which death will be relatively frequent.

Although the guidance documents do not specifically refer to a summary of SAEs, only a listing, we believe a summary table of all serious events during the treatment period is very helpful for data interpretation (Table 11.7). If the number of SAEs is expected to be extremely small, then a listing would be sufficient. The recommended summary table displays all PTs by decreasing frequency without sorting by SOC. The table could be sorted by SOC if deemed useful (eg, a lot of SAEs across a number of SOCs).

To understand further the severity of AEs, a summary of AEs leading to treatment discontinuation is recommended (Table 11.8). The guidance documents do not specifically refer to a summary table, only a listing, but it is commonly created across the industry and is considered very useful in understanding tolerability. The recommended summary table displays all preferred terms by decreasing frequency without sorting by SOC. The table could be sorted by SOC if deemed useful (eg, a lot of events leading to discontinuations across a number of SOCs).

Table 11.9 is recommended for AESIs that comprise a list of pre-defined preferred terms. Table 11.10 is recommended when there is an AESI that can be defined by the use of a SMQ. (The table would become more complex for SMQs with sub-SMQs.) Standardized MedDRA queries were created to define common medical concepts of interest by grouping relevant MedDRA PTs. See Section 10.4 for further discussion regarding SMQs.

Table 11.1.
Overview of Adverse Events During the Treatment Phase
Safety Population

	T1 (N=XX) n (%)	T2 (N=XX) n (%)	PL (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)
Treatment Emergent Adverse Events	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)
Serious Adverse Events	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)
Adverse Events leading to Discontinuation of Treatment	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)
Fatal Adverse Events	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.
Serious Adverse Events are counted regardless of whether or not severity increases relative to baseline.
aRisk difference confidence interval from [insert method].

Table 11.2.
All Deaths Safety Population

	T1 (N=XX) n (%)	T2 (N=XX) n (%)	PL (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)
All Deaths	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)
During the <<Treatment>> Phase	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)
During the <<Long Term Follow Up>> Phase	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 a row event.
Percentages are calculated relative to the treatment group N.
[a]Risk difference confidence interval from [insert method].

Table 11.3.
Summary of Treatment-Emergent Adverse Events by Preferred Term in Descending Frequency of T1 & T2 within System Organ Class Safety Population

System Organ Class Preferred Term	T1 (N=XX) n (%)	T2 (N=XX) n (%)	PL (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 treatment-emergent 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)
[System Organ Class #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 #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][a]	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)
[System Organ Class #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 #1][b]	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.
[a]Risk difference confidence interval from [insert method].
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].

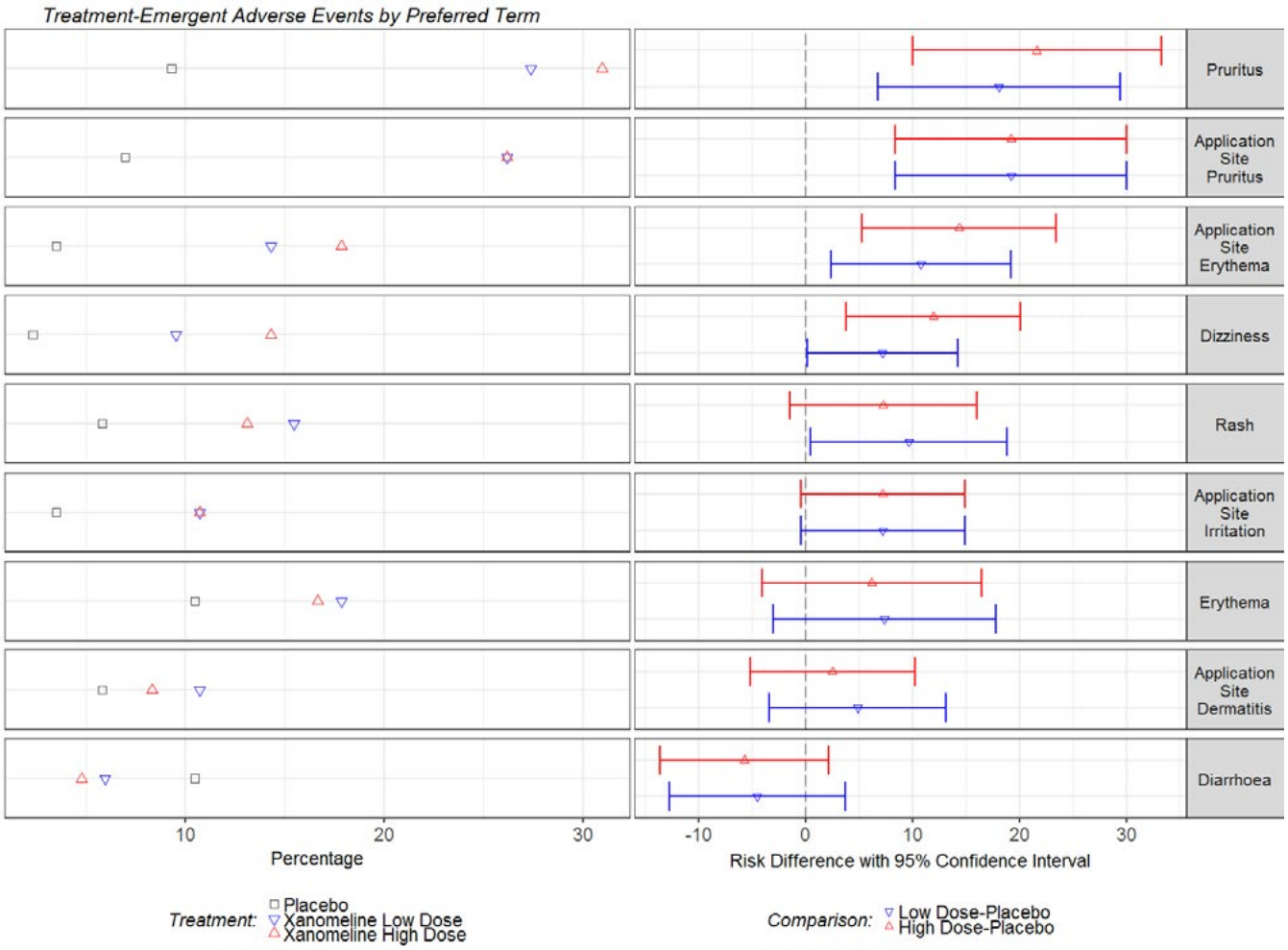
Table 11.4.
Summary of Common (≥1%) Treatment-Emergent Adverse Events by Preferred Term in Descending Frequency of T1 & T2 Safety Population

Preferred Term	T1 (N=XX) n (%)	T2 (N=XX) n (%)	PL (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)
[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][a]	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][b]	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.
Common is defined as an incidence of ≥1% in any group (T1, T2, or PL).

[a]Risk difference confidence interval from [insert method].
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].

Figure 11.1.
Common Treatment-Emergent Adverse Events



Adverse Events are based on 10% in either treatment group.

Table 11.5.
Summary of Treatment-Emergent Adverse Events by Maximum Severity and by Preferred Term in Descending Frequency of T1 & T2 within System Organ Class Safety Population

Preferred Term	Maximum Severity	T1 (N=XX) n (%)	T2 (N=XX) n (%)	PL (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 treatment-emergent adverse event	Mild	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Moderate	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Severe	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)
	Moderate/Severe	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)
	Any Severity	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)
[System Organ Class #1]	Mild	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Moderate	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Severe	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)
	Moderate/Severe	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)
	Any Severity	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]	Mild	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Moderate	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Severe	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)
	Moderate/Severe	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)
	Any Severity	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]	Mild	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Moderate	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Severe	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)
	Moderate/Severe	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)
	Any Severity	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.

[a]Risk difference confidence interval from [insert method].
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].

Table 11.6.
Summary of Treatment-Emergent Adverse Events by Maximum CTCAE Grade by Preferred Term in Descending Frequency of T1 & T2 within System Organ Class Safety Population

Preferred Term	Maximum Severity	T1 (N=XX) n (%)	T2 (N=XX) n (%)	PL (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 TEAE	Grade 1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Grade 2	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Grade 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)
	Grade 4	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)
	Grade 5	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)
	Grade > 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)
	Any Grade	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)
[System Organ Class #1]	Grade 1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Grade 2	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Grade 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)
	Grade 4	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)
	Grade 5	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)
	Grade > 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)
	Any Grade	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]	Grade 1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Grade 2	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Grade 3	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
	Grade 4	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)
	Grade 5	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)
	Grade > 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)
	Any Grade	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]	Grade 1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Grade 2	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	– –	– –	– –
	Grade 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)
	Grade 4	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)
	Grade 5	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)
	Grade > 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)
	Any Grade	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.

[a]Risk difference confidence interval from [insert method].
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].

Listing 11.1.
Listing of Deaths by Treatment Safety Population

Treatment = <<Treatment 1>>

Patient ID	Age	Sex	Race	Study Day of Last Study Drug Administration	Study Day of Death	Preferred Term of Fatal Event	Verbatim Term of Fatal Event

Table 11.7.
Summary of Serious Adverse Events by Preferred Term in Descending Frequency of T1 & T2 Safety Population

Preferred Term	T1 (N=XX) n (%)	T2 (N=XX) n (%)	PL (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 serious 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)

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.

Serious Adverse Events are counted regardless of whether severity increases relative to baseline.

[a]Risk difference confidence interval from [insert method].
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].

Table 11.8.
Summary of Adverse Events Leading to Treatment Discontinuation by Preferred Term in Descending Frequency of T1 & T2
Safety Population

Preferred Term	T1 (N=XX) n (%)	T2 (N=XX) n (%)	PL (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 adverse event leading to treatment discontinuation	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)

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.

[a]Risk difference confidence interval from [insert method].
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].

Table 11.9.
Summary of [Adverse Events of Special Interest] Defined Using a Pre-Defined List of MedDRA Preferred Terms by Preferred Term in Descending Frequency of T1 & T2
Safety Population

Preferred Term	T1 (N=XX) n (%)	T2 (N=XX) n (%)	PL (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.

[a]Risk difference confidence interval from [insert method].
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].

Table 11.10.
Summary of [Adverse Events of Special Interest] Defined Using Standardized MedDRA Query [SMQ Name] Safety Population

Preferred Term	T1 (N=XX) n (%)	T2 (N=XX) n (%)	PL (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)
Number of patients reporting at least one Narrow or Broad 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:
Broad and narrow scope preferred terms (PTs) are defined by a standardized MedDRA SMQ [or if not by MedDRA, provide appropriate detail, eg, sponsor defined custom search].

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].
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].

11.2. Discussion

There are certainly many ways to display AE data, and existing guidance documents vary in their recommendations and/or lack of detail. Hopefully the recommendations in this white paper will provide added details to facilitate consistent implementation in clinical study reports and integrated summary documents.

There was some debate as to whether an overview table (Table 11.1) should be recommended. The same information can be found dispersed across other recommended tables, so it can be viewed as superfluous. However, since the structure of a CSR and integrated summary document includes a brief overview of AEs, the overview table is recommended as something that can be included in the section. Otherwise, study teams may end up manually generating such a table, which is unnecessary if deemed of value during analysis planning.

The need for a summary of TEAEs is consistently highlighted in guidance documents. We assume at least one display that includes all TEAEs is needed. Since guidance documents consistently highlight the need for a display of "common" TEAEs, we also assume a display for common TEAEs is needed. Since the guidance documents specifically say displays sorted by SOC are desired (eg, the example table in the FDA Reviewer Guidance) [22], we assume at least one display sorted by SOC is needed. With all these assumptions in mind, and with the desire to not over-burden study teams with creating a bunch of similar tables with essentially the same information, we recommend a table of all TEAEs sorted by SOC (Table 11.3) and a figure (for all except very small studies) of common TEAEs not sorted by SOC (Figure 11.1). Creating a figure instead of a table for common TEAEs using MedDRA PTs was debated as it may give too much emphasis on a display that does not incorporate consolidation. (See Section 10.4 for further discussion on consolidation.) For example, abdominal pain may have been reported at high enough frequency to be considered common. However, "abdominal pain upper" and "abdominal pain lower" may not have been reported at high enough frequency and will not even be seen in a table or figure of common TEAEs making a medical review of such a display have limited value. While this is an issue for either a table or a figure, some feel creating a figure gives too much emphasis to a display with limited value. In general, despite this limitation, we recommend the figure except when the study is very small or when consolidation is of particular concern for the compound. Figure 17.3 was considered, but Figure 11.1 was chosen as the preferred display. The volcano plot in Ma and colleagues [14] could be a useful plot especially among compound teams that find p-values useful, and the figures in Southworth and O'Connell [52] could be worth exploring. We plan to discuss these plots and potentially other plots for future versions of this white paper.

We considered a table incorporating the MedDRA HLT to at least partially address the potential over-granularity of the MedDRA PT. We decided that the inclusion of HLTs or other ways to group terms (other than the MedDRA PT and SOC) is best done using interactive tools (see Section 10.4 and Section 18, Appendix B) when analyzing individual studies. For integrated summaries, we do recommend a more robust set of tables to address potential over-granularity (see Section 12.1). If a compound team does not have mature interactive tool utilization or another method to address potential over-granularity issues, this table should be considered. Future work is needed to more efficiently address grouping of terms to meet the needs of signal

detection and reporting frequencies in scientific disclosures (eg, cross-industry development of MedDRA Labeling Groupings). We plan to continue discussion on which displays tend to be most helpful for future versions of this white paper.

We considered a table that provides a summary of MedDRA PTs not sorted by HLT or SOC (Table 17.1, which has the entire list of MedDRA PTs; Table 11.4 or Figure 11.1 with common events not sorted by HLT or SOC are part of the recommendations). While this table is desired by some users, we decided it doesn't add sufficient value over the already existing tables and figures within the recommendations.

There was a lot of discussion related to a display of TEAEs considered related to study drug by the investigator. The exact same display recommended for all TEAEs was strongly considered (with the added requirement to only include TEAEs considered related to study drug by the investigator in the table; Table 17.2), as it would be the easiest to implement and it reflects current practice across the industry. As noted in Section 10.6, we recommend using the data for individual case reviews and/or provide the ability to review relatedness data into interactive display tools for exploration purposes instead of creating static tables. However, a summary is likely needed for submissions to the PMDA. See Section 12.1 for the recommended integrated display for submissions to the PMDA.

Regarding the display of TEAEs by severity (or CTCAE grade), a figure was considered (Figure 17.2). Although we are trying to move more toward visual displays, we decided that a table was better when displaying severity levels for all TEAEs. The figure can be utilized for AESIs if deemed useful. Additionally, there was some debate whether to display severity level for all TEAEs or just the common TEAEs. In the FDA Clinical Review Template [8] severity is discussed in the context of common events. In the EU Guidance [17], severity for all events is requested. Therefore, we recommend the table by severity levels to be created for all events, even though the table will be generally quite large. As noted in Section 10.5, we believe severity might be best reviewed utilizing interactive display tools but will defer this to future discussions. For now, the table could be created and made available upon request if it was not used for any key conclusions (instead of automatically including it in CSRs).

Regarding the display of SAEs (Table 11.7), there was discussion whether all SAEs during the treatment period versus "treatment-emergent" SAEs should be displayed. For some companies, "during the treatment period" and "treatment emergent" are equivalent concepts making the discussion moot. However, for other companies, the concept is not equivalent. For example, for companies that define treatment emergence as requiring a higher severity level during treatment compared to baseline, it is theoretically possible for an event to become serious during the treatment period (eg, hospitalized due to the event) and not be "treatment-emergent" (same severity level during the treatment period compared to baseline). To illustrate, let us say a patient has moderate depression prior to treatment and continues to have moderate depression throughout the study. At some point during the treatment period, the physician decides it has been long enough and the patient should be hospitalized making it "serious". In data collection, severity remained as moderate. In these cases, we believe it is better to count all events that become serious during the treatment period regardless of severity level and regardless how treatment-emergence is defined for AEs.

As noted in Section 10.9, there are circumstances where the exposure-adjusted incidence rate would be of interest. Table 17.3 provides an example if warranted. In addition, as discussed in Section 10.9, there are circumstances where a Kaplan-Meier plot would be useful. Figure 17.3 provides an example.

12. Tables and Figures for Integrated Summaries

12.1. 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 10.1). 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 randomization 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, if a sponsor decides not to utilize confidence intervals as a tool for review, they can be deleted. In addition, if a sponsor decides to utilize p-values as a tool for review, they can be added. See Table 12.1 through Table 12.10 and Figure 12.1.

As noted in Secton 11.2, creating a display of severity levels for all TEAEs (Table 12.5 and Table 12.6) is requested by the EMA [17]. Because the table would be quite large, sponsors can consider making it available upon request if it was not used for any key conclusions (instead of automatically including it in submissions).

Generally, analyses of safety data by subgroups (eg, sex, age, race) are not conducted for individual studies and are reserved for integrated summaries (unless a study is very large). We recommend Figure 12.2, a visual display for each common TEAE. The same display can be used for AESIs if there are enough cases. Additional subgroups (eg, weight groups) can be added given the subgroups of interest for the compound. Alternative ways of reviewing data by subgroups (eg, showing all events for a particular subgroup) can best be implemented in interactive display tools.

Table 12.11 is recommended when there are additional events that should be grouped together for signal detection and scientific disclosure (eg, labeling) purposes. As noted in Section 10.4, we recommend a table that incorporates the MedDRA HLT for integrated summaries partially to address potential over-granularity issues. However, we additionally recommend this table to include groupings not already addressed with the MedDRA HLT when applicable. Table 12.12 is recommended based on expectations from the EU [17]. Table 12.13 is recommended based on expectations from the PMDA [20] and is recommended for inclusion for submissions to the PMDA only.

Table 12.1.
Overview of Adverse Events During the Treatment Phase: Integrated Database Safety Population

	TRT (N=XX)		PL (N=XX)		
	n (%)	Study Size-Adjusted %	n (%)	Study Size-Adjusted %	Risk Difference[a] TRT-PL (95% CI)
Treatment-Emergent Adverse Events	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
Serious Adverse Events Serious Adverse Events	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
Adverse Events leading to Discontinuation of Investigational Product	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
Fatal Adverse Events	xx (xx.x)	xx.x	xx (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.

Serious Adverse Events are counted regardless of whether severity increases relative to baseline.

[a]Mantel-Haenszel risk difference stratified by study.

Table 12.2.
Summary of Treatment-Emergent Adverse Events by Preferred Term in Descending Frequency of TRT within System Organ Class: Integrated Database Safety Population

System Organ Class Preferred Term	TRT (N=XX)		PL (N=XX)		
	n (%)	Study Size-Adjusted %	n (%)	Study Size-Adjusted %	Risk Difference[a] TRT-PL (95% CI)
Number of patients reporting at least one treatment-emergent adverse event	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[System Organ Class #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #2][a]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[System Organ Class #2]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #1][b]	xx (xx.x)	xx.x	xx (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.

[a]Mantel-Haenszel risk difference stratified by study.
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].

Table 12.3.
Summary of Treatment-Emergent Adverse Events by High Level Term and Preferred Term in Descending Frequency of TRT within System Organ Class
Safety Population

System Organ Class High Level Term Preferred Term	TRT (N=XX)		PL (N=XX)		
	n (%)	Study Size-Adjusted %	n (%)	Study Size-Adjusted %	Risk Difference[a] TRT-PL (95% CI)
Number of patients reporting at least one treatment-emergent adverse event	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[System Organ Class #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[High Level Term #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #2][a]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[High Level Term #2]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #1][b]	xx (xx.x)	xx.x	xx (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.

[a]Mantel-Haenszel risk difference stratified by study.
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].

Table 12.4.
Summary of Common (≥1%) Treatment-Emergent Adverse Events by Preferred Term in Descending Frequency: Integrated Database
Safety Population

System Organ Class High Level Term Preferred Term	TRT (N=XX)		PL (N=XX)		
	n (%)	Study Size-Adjusted %	n (%)	Study Size-Adjusted %	Risk Difference[a] TRT-PL (95% CI)
[Preferred Term #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #2][a]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #3][b]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)

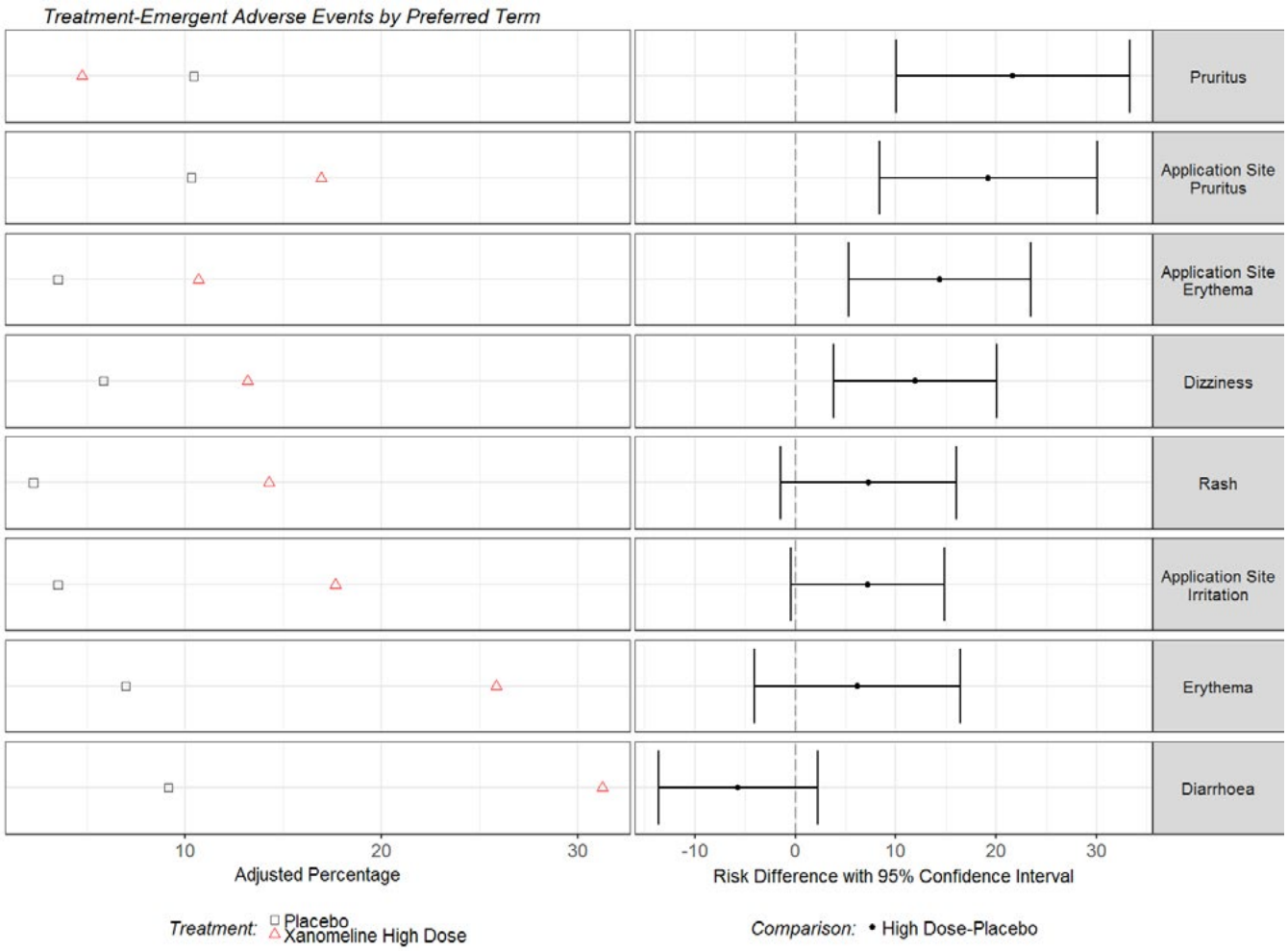
Footnotes:
N=number of patients; n=number of patients with at least one row event. Percentages are calculated relative to the treatment group N.

Patients may be counted in more than one row.

Common is defined as an incidence of ≥1% in any group (TRT or PL).

[a]Mantel Haenszel risk difference stratified by study.-
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX(Placebo) [For applicable PTs].

Figure 12.1.
Common Treatment-Emergent Adverse Events



Adverse Events are based on 10% in either treatment group.

Table 12.5.
Summary of Treatment-Emergent Adverse Events by Maximum Severity by Preferred Term in Descending Frequency of TRT within System Organ Class: Integrated Database
Safety Population

Preferred Term	Maximum Severity	TRT (N=XX)		PL (N=XX)		
		n (%)	Study Size-Adjusted %	n (%)	Study Size-Adjusted %	Risk Difference[a] TRT-PL (95% CI)
Number of patients reporting at least one treatment-emergent adverse event	Mild	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Moderate	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Severe	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Moderate/Severe	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Any Severity	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[System Organ Class #1]	Mild	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Moderate	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Severe	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Moderate/Severe	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Any Severity	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #1]	Mild	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Moderate	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Severe	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Moderate/Severe	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Any Severity	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #2]	Mild	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Moderate	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Severe	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Moderate/Severe	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Any Severity	xx (xx.x)	xx.x	xx (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.

[a]Mantel Haenszel risk difference stratified by study.-
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX(Placebo) [For applicable PTs].

Table 12.6.
Summary of Treatment-Emergent Adverse Events by Maximum CTCAE Grade by Preferred Term in Descending Frequency of TRT within System Organ Class: Integrated Database
Safety Population

Preferred Term	Maximum CTCAE Grade	TRT (N=XX)		PL (N=XX)		
		n (%)	Study Size-Adjusted %	n (%)	Study Size-Adjusted %	Risk Difference[a] TRT-PL (95% CI)
Number of patients reporting at least one treatment-emergent adverse event	Grade 1	xx (xx.x)	xx.x	xx (xx.x)	xx.x	– –
	Grade 2	xx (xx.x)	xx.x	xx (xx.x)	xx.x	– –
	Grade 3	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Grade 4	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Grade 5	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Grade > 3	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Any Grade	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[System Organ Class #1]	Grade 1	xx (xx.x)	xx.x	xx (xx.x)	xx.x	– –
	Grade 2	xx (xx.x)	xx.x	xx (xx.x)	xx.x	– –
	Grade 3	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Grade 4	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Grade 5	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Grade > 3	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Any Grade	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #1]	Grade 1	xx (xx.x)	xx.x	xx (xx.x)	xx.x	– –
	Grade 2	xx (xx.x)	xx.x	xx (xx.x)	xx.x	– –
	Grade 3	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Grade 4	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Grade 5	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Grade > 3	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Any Grade	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #2]	Grade 1	xx (xx.x)	xx.x	xx (xx.x)	xx.x	– –
	Grade 2	xx (xx.x)	xx.x	xx (xx.x)	xx.x	– –
	Grade 3	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Grade 4	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Grade 5	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Grade > 3	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
	Any Grade	xx (xx.x)	xx.x	xx (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.

[a]Mantel-Haenszel risk difference stratified by study.
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX(Placebo) [For applicable PTs].

Listing 12.1.
Listing of Deaths by Study and Treatment: Integrated Database
Safety Population
Study XXXX

Treatment = <<Treatment 1>>							
Patient ID	**Age**	**Sex**	**Race**	**Study Day of Last Study Drug Administration**	**Study Day of Death**	**Preferred Term of Fatal Event**	**Verbatim Term of Fatal Event**

Table 12.7.
Summary of Serious Adverse Events by Preferred Term in Descending Frequency of TRT: Integrated Database
Safety Population

Preferred Term	TRT (N=XX)		PL (N=XX)		
	n (%)	Study Size-Adjusted %	n (%)	Study Size-Adjusted %	Risk Difference[a] TRT-PL (95% CI)
Number of patients reporting at least one serious adverse event	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #2]	xx (xx.x)	xx.x	xx (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.

Serious Adverse Events are counted regardless of whether severity increases relative to baseline.

[a]Mantel-Haenszel risk difference stratified by study.
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX(Placebo) [For applicable PTs].

Table 12.8.
Summary of Adverse Events Leading to Treatment Discontinuation by Preferred Term in Descending Frequency of TRT: Integrated Database
Safety Population

Preferred Term	TRT (N=XX)		PL (N=XX)		
	n (%)	Study Size-Adjusted %	n (%)	Study Size-Adjusted %	Risk Difference[a] TRT-PL (95% CI)
Number of patients reporting at least one adverse event leading to treatment discontinuation	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #2]	xx (xx.x)	xx.x	xx (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.

Serious Adverse Events are counted regardless of whether severity increases relative to baseline.

[a]Mantel-Haenszel risk difference stratified by study.
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX(Placebo) [For applicable PTs].

Table 12.8.
Summary of [Adverse Events of Special Interest] Based on a Pre-Defined List of MedDRA Preferred Terms by Preferred Term in Descending Frequency of TRT: Integrated Database
Safety Population

Preferred Term	TRT (N=XX)		PL (N=XX)		
	n (%)	Study Size-Adjusted %	n (%)	Study Size-Adjusted %	Risk Difference[a] TRT-PL (95% CI)
Number of patients reporting at least one [Event Cluster] adverse event	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #2]	xx (xx.x)	xx.x	xx (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.

[a]Mantel-Haenszel risk difference stratified by study.
[a]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicablebPTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX(Placebo) [For applicable PTs].[For applicable PTs].

Table 12.10.
Summary of [Adverse Events of Special Interest] Based on Standardized MedDRA Query [SMQ]: Integrated Database Safety Population

Preferred Term	TRT (N=XX)		PL (N=XX)		
	n (%)	Study Size-Adjusted %	n (%)	Study Size-Adjusted %	Risk Difference[a] TRT-PL (95% CI)
Number of patients reporting at least one Narrow Scope PT	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
Number of patients reporting at least one Narrow or Broad Scope PT	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
Narrow Scope PTs					
[Preferred Term #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #2]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #n]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
Broad Scope PTs					
[Preferred Term #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #2]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #n]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)

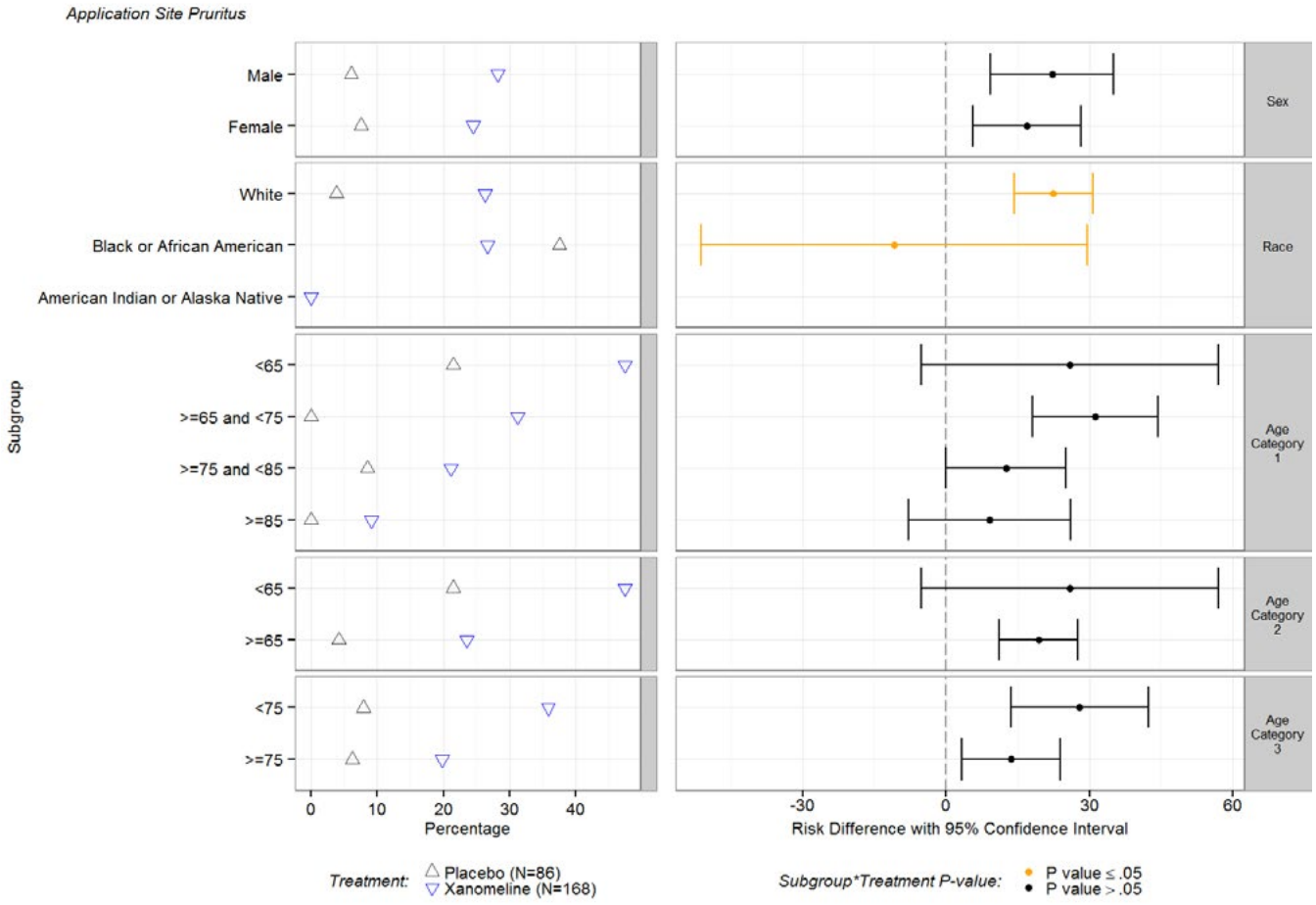
Footnotes:
N=number of patients; n=number of patients with at least one row event. Percentages are calculated relative to the treatment group N.

Patients may be counted in more than one row.

Broad and narrow scope preferred terms (PTs) are defined by a standardized MedDRA SMQ [or if not by MedDRA, provide appropriate detail, eg, sponsor defined custom search].

[a]Mantel-Haenszel risk difference stratified by study.
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX(Placebo) [For applicable PTs].

Figure 12.2.
Treatment-Emergent Adverse Events by Subgroup



P value is from the test of subgroup*treatment interaction in the logistic regression model.
The model includes treatment, subgroup and treatment*subgroup as independent variables.

Table 12.11.
Summary of Consolidated Events: Integrated Database Safety Population

Preferred Term	TRT (N=XX)		PL (N=XX)		
	n (%)	Study Size-Adjusted %	n (%)	Study Size-Adjusted %	Risk Difference[a] TRT-PL (95% CI)
Number of patients reporting at least one [Event Cluster 1] adverse event	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #2]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
Number of patients reporting at least one [Event Cluster 2] adverse event	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx.x (xx.x, xx.x)
[Preferred Term #2]	xx (xx.x)	xx.x	xx (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.

[a]Mantel-Haenszel risk difference stratified by study.
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX(Placebo) [For applicable PTs].

Table 12.12.
Overview of Adverse Events by Age Category
All Patients Treated with Investigational Product

Preferred Term	Age: < 65 years (N=XX)	Age: 65-74 years (N=XX)	Age: 75-84 years (N=XX)	Age: ≥ 85 years (N=XX)
	n (%)	n (%)	n (%)	n (%)
Total TEAEs	xxx (xx.x)	xxx (xx.x)	xxx (xx.x)	xxx (xx.x)
SAEs	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Fatal	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Hospitalization	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Life-Threatening	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Disability	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Other	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
AEs leading to study drug discontinuation	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Accidents and injuries (SMQ)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Cardiac disorders (SOC)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Infections and infestations (SOC)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Nervous system disorders (SOC)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Psychiatric disorders	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Vascular disorders	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
Hypotension, falls, fractures[a]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
QOL decreased (PT)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)
[Add other topics per compound]	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (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.

[a]Hypotension, falls, fractions is defined by: Any MedDRA PT that is included in the "Decreased and nonspecific blood pressure disorders and shock" HLGT, the "Fractures and dislocations NEC" HLT, the "Limb fractures and dislocations" HLT, the "Pelvic fractures and dislocations" HLT, the "Skull fractures, facial bone fractures and dislocations" HLT, the "Spinal fractures and dislocations" HLT, the "Thoraciccage fractures and dislocations" HLT, or the "Fall" PT.

Table 12.13.
Summary of Treatment-Emergent Adverse Events (All-Causality and Those Considered Related by the Investigator) by Preferred Term in Descending Frequency with System Organ Class Integrated Safety Population

	All-Causality				Treatment-Related by Investigator			
Preferred Term	TRT (N=XX)		PL (N=XX)		TRT (N=XX)		PL (N=XX)	
	n (%)	Study Size-Adjusted %	n (%)	Study Size-Adjusted %	n (%)	Study Size-Adjusted %	n (%)	Study Size-Adjusted %
Number of patients reporting at least one treatment-emergent adverse event	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx (xx.x)	xx.x
[System Organ Class #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx (xx.x)	xx.x
[Preferred Term #1]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx (xx.x)	xx.x
[Preferred Term #2][a]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx (xx.x)	xx.x
[System Organ Class #2]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx (xx.x)	xx.x
[Preferred Term #1][b]	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx (xx.x)	xx.x	xx (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.

[a]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[b]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].

12.2. Discussion

The discussion in Section 11.2 also applies to integrated summaries and will not be repeated.

Regarding the figure for subgroups, there was discussion on whether it is better to show all the subgroups for each common event (Figure 12.2) or display all the common events for a particular subgroup (Figure 17.4). As noted in Section 12.1, we believe interactive display tools can be leveraged to allow medical to see data for subgroups in various formats. For submission purposes, we believe the figure that shows all the subgroups for each common event is the most user-friendly and allows for consistency on how data is displayed between common TEAEs and AESIs. For subgroups, tables were also considered (Table 17.4 and Table 17.5); however, figures are recommended since they really facilitate the review in determining any potential differences in the AE profile across subgroups.

13. Example Statistical Analysis Plan Language

13.1. Individual Study

The planned analyses of adverse events (AEs) are based on the recommendations provided in a white paper produced by a Pharmaceuticals User Software Exchange (PHUSE) Computational Science Working Group (a collaboration with the Food and Drug Administration [FDA] and PHUSE). The white paper includes justifications for the choices. Specifically, the following white paper pertains to AEs: Analysis and Displays Associated with Adverse Events: Focus on Adverse Events in Phase 2-4 Clinical Trials and Integrated Submission Documents (*http://www.phuse.eu/CSS-deliverables.aspx*).

Not all displays described in this plan will necessarily be included in the clinical study report. Not all displays will necessarily be created as a "static" display. Some may be incorporated into interactive display tools instead of or in addition to a static display. Any display described in this plan and not provided would be available upon request.

A treatment-emergent adverse event (TEAE) will be defined using the "LLT method" in Crowe and colleagues [10]. Specifically, the Medical Dictionary for Regulatory Activities (MedDRA) lowest level term (LLT) will be used in the treatment-emergent computation. The maximum severity for each LLT during the baseline period will be used as baseline. The treatment period will be included as post-baseline for the analysis. An event will be considered treatment-emergent if the maximum severity during the treatment period is greater than the maximum severity during baseline. For events that are sex-specific, the denominator and computation of the percentage will include only patients from the given sex.

For some displays (as noted), the risk difference and 95% confidence interval will be displayed for each treatment pair (high dose minus placebo, low dose minus placebo, high dose minus low dose) for each MedDRA preferred term.

In an overview table, the number and percentage of patients who experienced a TEAE, serious adverse event (SAE), died due to an AE, or discontinued from study treatment due to an AE will be summarized by treatment. Risk differences and 95% confidence intervals will also be displayed.

The percentages of patients with TEAEs will be summarized by treatment using MedDRA preferred term nested within system organ class (SOC). Events will be ordered by decreasing frequency within SOC, and SOC with be ordered by decreasing frequency. Risk differences and 95% confidence intervals will

also be displayed. Interactive tools will be used to review the data using additional levels of MedDRA (eg, high level term, high level group term), and/or different ways to group terms to assist in medical review.

In a figure, the percentages of patients with TEAEs and risk differences with 95% confidence intervals will be summarized by treatment using MedDRA preferred term for the common TEAEs (occurred in 1% before rounding of treated patients). Events will be ordered by risk difference (high dose minus placebo).
The percentages of patients with TEAEs by maximum severity will be summarized by treatment using MedDRA preferred term for the common TEAEs. For each patient and TEAE, the maximum severity for the MedDRA preferred term is the maximum post-baseline severity observed from all associated LLTs mapping to the MedDRA preferred term. Risk differences and 95% confidence intervals will be included for severe events and moderate/severe events combined.

A listing will be produced that includes all deaths from the study. All deaths will be included, regardless of the investigator's or the sponsor's judgment about causality.

The number and percentage of patients who experienced a SAE (including the SAEs that led to death) during the treatment period (without regard to severity levels during treatment relevant to severity levels during baseline) will be summarized by treatment using MedDRA preferred term. Events will be ordered by decreasing frequency. Risk differences and 95% confidence intervals will also be displayed.

The number and percentage of patients who permanently discontinued from study treatment due to an AE (including AEs that led to death) during the treatment period will be summarized by treatment using MedDRA preferred term. Events will be ordered by decreasing frequency. Risk differences and 95% confidence intervals will also be displayed.

13.2. Integrated Summary

The planned integrated analyses of adverse events (AEs) are based on the recommendations provided in a white paper produced by a Pharmaceuticals User Software Exchange (PHUSE) Computational Science Working Group (a collaboration with the Food and Drug Administration [FDA] and PHUSE). The white paper includes justifications for the choices. Specifically, the following white paper pertains to AEs: Analysis and Displays Associated with Adverse Events: Focus on Adverse Events in Phase 2-4 Clinical Trials and Integrated Submission Documents (*http://www.phuse.eu/CSS-deliverables.aspx*).

Determing which AEs are adverse drug reactions (ADRs) will be based on the review of all the planned summaries and/or analyses, plus any additional ad-hoc displays and case reviews needed to assist the team in the determination. Factors used to determine the list of ADRs include the following: biologic plausibility, clinical impressiveness of any individual case (eg, any available dechallenge/rechallenge information), the statistical assessment of the strength and magnitude of the observed effect via risk differences and confidence intervals, the observed dose relationship, the severity of the event, the consistency of findings across studies, the consistency of findings from similar events, and the consistency of findings from similar compounds, and clinical relevance.

Not all displays described in this plan will necessarily be included in the Summary of Clinical Safety. Not all displays will necessarily be created as a "static" display. Some may be incorporated into interactive display tools instead of or in addition to a static display. Any display described in this plan and not provided would be available upon request. A treatment-emergent adverse event (TEAE) will be defined using the "LLT method" in Crowe and colleagues [10]. Specifically, the Medical Dictionary for Regulatory Activities (MedDRA) lowest level term (LLT) will be used in the treatment-emergent computation. The maximum severity for each LLT during the baseline period will be used as baseline. The treatment period will be included as post-baseline for the analysis. An event will be considered treatment-emergent if the maximum severity during the treatment period is greater than the maximum severity during baseline. For events that are sex-specific, the denominator and computation of the percentage will include only patients from the given sex. All doses of the investigational drug that fall within the range of draft label dosing will be included as a single treatment arm.

In some displays (as noted), the Mantel-Haenszel risk difference stratified by study and 95% confidence interval will be displayed (treatment minus placebo). Study-size adjusted percentages will also be presented. See Crowe and colleagues [13] for methodology.

In an overview table, the number and percentage of patients who experienced a TEAE, serious adverse event (SAE), died due to an AE, or discontinued from study treatment due to an AE will be summarized by treatment. Mantel-Haenszel risk differences and 95% confidence intervals, and study-size adjusted percentages will also be displayed.

The percentages of patients with TEAEs will be summarized by treatment using MedDRA preferred term nested within system organ class (SOC). Events will be ordered by decreasing frequency within SOC. As an additional table, the percentages of patients with TEAEs will be summarized by treatment using MedDRA preferred term nested within MedDRA high level term (HLT) and SOC. Events will be ordered by decreasing frequency based on the MedDRA HLT, and SOC will be ordered by decreasing frequency. Mantel-Haenszel risk differences and 95% confidence intervals, and study-size adjusted percentages will also be displayed. Interactive tools will also be used to review the data using different ways to group terms to assist in medical review.

In a figure, the percentages of patients with TEAEs and risk differences with 95% confidence intervals will be summarized by treatment using MedDRA preferred term for the common TEAEs (occurred in 1% before rounding of treated patients). Events will be ordered by risk difference (treatment minus placebo).

The percentages of patients with TEAEs by maximum severity will be summarized by treatment using MedDRA preferred term for the common TEAEs. For each patient and TEAE, the maximum severity for the MedDRA preferred term is the maximum post-baseline severity observed from all associated LLTs mapping to the MedDRA preferred term. Mantel-Haenszel risk differences and 95% confidence intervals, and study-size adjusted percentages will also be displayed for severe events and moderate/severe events combined.

A listing will be produced that includes all deaths from all

studies comprising the submission. All deaths will be included, regardless of the investigator's or the sponsor's judgment about causality.

The number and percentage of patients who experienced a SAE (including the SAEs that led to death) during the treatment period (without regard to severity levels during treatment relevant to severity levels during baseline) will be summarized by treatment using MedDRA preferred term. Events will be ordered by decreasing frequency. Mantel-Haenszel risk differences and 95% confidence intervals, and study-size adjusted percentages will also be displayed.

The number and percentage of patients who permanently discontinued from study treatment due to an AE (including AEs that led to death) during the treatment period will be summarized by treatment using MedDRA preferred term. Events will be ordered by decreasing frequency. Mantel-Haenszel risk differences and 95% confidence intervals, and study-size adjusted percentages will also be displayed.

Summary tables comparing exposure-adjusted incidence rates over the time during the treatment period will be generated for the common TEAEs. The exposure-adjusted incidence rate will be calculated as the total number of patients who experienced the TEAE for each MedDRA preferred term divided by the total time at risk during the time interval. Time at risk will be calculated as the sum of time to first event for patients who experienced the TEAE and the time during the interval for patients who do not experience the TEAE. These tables will include the total number of patients in each group, the total person years on study, the total person-time at risk per event, the incidence rate (number of patients with the event / total time at risk), incidence rate ratio and 95% confidence interval. Both the incidence rate ratio and 95% confidence interval will be derived from a Poisson regression model with study and treatment as explanatory variables.

Figures displaying analyses for subgroups will be created for each common TEAE. The following are the subgroups that will be analyzed:

- age group 1 (<65, ≥65 to <75, ≥75 to <85 and ≥85)
- age group 2 (<65 vs ≥65)
- age group 3 (<75 vs ≥75)
- sex
- race

Treatment-emergent adverse events will be presented by MedDRA preferred term. A logistic regression model using the Firth correction [53] will be used to test the treatment-by-subgroup interaction. Interaction tests with p-values less than 0.05 will be colored. The response variable will be each most common TEAE. The explanatory variables will be study, treatment, subgroup, and treatment-by-subgroup interaction. Within each subgroup category, the risk difference and 95% confidence interval will be presented in the figure. The logistic regression model will exclude subgroups in the Missing category and those with fewer than 5 patients (not 5 patients with events) in either of the treatment arms. The condition on 5 patients is needed to enable the logistic regression to converge. If required, the number of patients to enable convergence will be increased (in increments of 5). Note that the test of interaction is testing for a difference in odds ratios among subgroups, whereas the within-subgroup estimate and confidence intervals are based on the risk difference. Thus, it is possible that there will be a disconnect between what is shown on the figure versus what has been colored as having an interaction test with a p-value less than 0.05.

[Note: Displays for dose relationship are not covered in this white paper, but provided as example statistical anlaysis plan language to show a possible method for assessing dose relationship.]

In addition, dose relationship will be assessed for the common TEAEs. Studies with fixed doses of investigational study drug and placebo will be included. Since not all studies have the same fixed doses, an indirect Bayesian method will be used and applied to each common MedDRA preferred term. The model to be used is the "Linear Logistic Indirect Comparison (LLIC) Model" [54]. Of interest, the posterior mean of the 'dose response' parameter indicates the magnitude of relationship that the event has to an increase in one increment of dose. The posterior standard deviation associated with that parameter estimate is given. The Monte Carlo error is the difference between the mean of the sampled values (which we are using as our estimate of the posterior mean for each parameter) and the true posterior mean (reference WinBUGS user manual, *http://www.mrc-bsu.cam.ac.uk/bugs*). To evaluate the probability that a dose relationship exists, the 95% credible interval limits are provided along with the median. If the lower 2.5% limit is greater than 0, a dose relationship is considered to be likely. The "Prob a ≥ 0" provides the posterior probability that the true parameter is greater than 0.

14. Acknowledgements

The primary contributors include Mary Nilsson, Sheryl Treichel, Nhi Beasley, Brenda Crowe, Robert Temple, Rebeka Revis, and Wei Wang.

Additional contributors/reviewers (eg, participated in discussions, provided review comments on portions of the content) include: Sascha Ahrweiler, Greg Anglin, Greg Ball, Timothy Barnett, Simin Baygani, Charles Beasley, Coen Bernaards, Gustav Bernard, Cathy Bezek, Nancy Brucken, Alan Cantrell, Asa Carlsheimer, Sunita Dhar, Susan DeHaven, Mary Anne Dellva, Janet Fan, Ross Farrugia, Josephine Fong, Jim Gaiser, Abigail Garchitorena, Sterling Hardy, Weizhong He, Claire Hughes, Karolyn Kracht, Karin LaPann, Misun Lee, Kathryn Li, Grace Lu, Kristen Harrington, Lisa Houterloot, Jane Lozano, Lachlan MacGregor, Vidya Maiya, Rich Manski, Yoshihiro Nakashima, Mercidita Navarro, Russ Newhouse, Pierre Nicolas, Casie Polanco, Mithun Ranga, Douglas Roepke, Dustin Ruff, Anne Russotto, Frank Senk, Julie Shah, Jun Takeda, Yusuf Tanrikulu, Lila Thome, Angelo Tinazzi, Raul Vinueza, Terry Walsh, Michael Ward, and Xiaoping Zhang. Apologies to contributors/reviewers that we may have missed.

15. Project Leader Contact Information

Name: Mary Nilsson (Analysis and Display White Papers Project Team leader) Enterprise: Eli Lilly & Company City, State ZIP: Indianapolis, IN 46285 Work Phone: 317-651-8041
E-mail: *nilsson_mary_e@lilly.com*

16. References

All links in this section were accessed December 14, 2016.

1. CIOMS Working Group VI. Management of safety information from clinical trials. 2005, Geneva: Council for International Organizations of Medical Sciences
2. A dictionary of epidemiology. 6th ed. 2014, New York, NY: Oxford University Press.
3. A dictionary of epidemiology. 4th ed. 2001, New York, NY: Oxford University Press.
4. Zhou Y, Ke C, Jiang Q, Shahin S, Snapinn S. Choosing appropriate metrics to evaluate adverse events in safety evaluation. Therapeutic Innovation & Regulatory Science, 2015: 2168479014565470.
5. U. S. Department of Health and Human Services. Morbidity frequency measures (Lesson 3, Section 2), in Principles of epidemiology in public health practice, third edition: An introduction to applied epidemiology and biostatistics. 2006, Centers for Disease Control and Prevention: Atlanta, GA.
6. Rothman K, Greenland S, Lash T. Modern epidemiology. 3rd Edition ed. 2008, Philadelphia, PA: Lippincott, Williams & Wilkins.
7. Handbook of epidemiology. 2005, Berlin, Heidelberg, New York: Springer-Verlag Berlin Heidelberg.
8. U.S. Food and Drug Administration. Good review practice: Clinical review template. Manual of policies and procedures, 2010. *http://www.fda.gov/downloads/AboutFDA/CentersOffices/CDER/ManualofPoliciesProcedures/ucm080121.pdf*.
9. U.S. Food and Drug Administration. Guidance for industry and investigators: Safety reporting requirements for INDs and BA/BE studies. 2012, Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research.
10. Crowe B, Brueckner A, Beasley C, Kulkarni P. Current practices, challenges, and statistical issues with product safety labeling. Stat Biopharm Res, 2013, 5(3): 180-193.
11. Crowe B, Xia HA, Berlin J, Watson DJ, Shi H, Lin SL, Kuebler J, Schriver RC, Santanello N, Rochester G, Porter JB, Oster M, Mehrotra D, Li Z, King EC, Harpur ES, Hall DB. 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. Clinical Trials, 2009, 6(5): 430-440.
12. Chuang-Stein C, Beltangady M. Reporting cumulative proportion of subjects with an adverse event based on data from multiple studies. Pharmaceutical Statistics, 2011, 10(1): 3-7.
13. Crowe B, Chuang-Stein C, Lettis S, Brueckner A. Reporting adverse drug reactions in product labels. Therapeutic Innovation & Regulatory Science, 2016, 50(4): 455-463.
14. Ma H, Ke C, Jiang Q, Snapinn S. Statistical considerations on the evaluation of imbalances of adverse events in randomized clinical trials. Therapeutic Innovation & Regulatory Science, 2015, 49(6): 957-965.
15. Rosario LA, Kropp TJ, Wilson SE, Cooper CK. Join FDA/PHUSE working groups to help harness the power of computational science. Drug Information Journal, 2012, 46(5): 523-524.
16. U.S. Food and Drug Administration. Guidance for industry: Determining the extent of safety data collection needed in late-stage premarket and postapproval clinical investigations. 2016. *http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm291158.pdf*.
17. European Medicines Agency. Guidance document for the content of the <Co->Rapporteur day 80 critical assessment report. 2014. *http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2009/10/WC500004856.pdf*.
18. U.S. Food and Drug Administration. Guidance for industry: E3 structure and content of clinical study reports. Questions and answers (R1), 2013. *http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm336889.pdf*.
19. CDISC Analysis Data Model Team. Analysis data model (ADaM) data structure for adverse event analysis. 2012. *https://www.cdisc.org/sites/default/files/members/standard/foundational/adam/adam_ae_final_v1.pdf*.
20. Japanese Administrative Notice. Format for preparing the common technical document for submission of New Drug Applications to reduce total review time. 2011. *http://home.att.ne.jp/red/akihiro/mhlw/CTD_20110322.pdf*.
21. European Commission; Enterprise and Industry Directorate-General. A guideline on Summary Of Product Characteristics (SmPC). 2009. *http://ec.europa.eu/health/files/eudralex/vol-2/c/smpc_guideline_rev2_en.pdf*.
22. U.S. Food and Drug Administration. Reviewer guidance: Conducting a clinical safety review on a new product application and preparing a report on the review. 2005. *http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072974.pdf*.
23. U.S. Food and Drug Administration. Guidance for Industry: Premarketing risk assessment. 2005. *http://www.fda.gov/downloads/RegulatoryInformation/Guidances/UCM126958.pdf*.
24. International Conference on Harmonisation Expert Working Group. 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 Reports, 2002.
25. International Conference on Harmonization. ICH E9 statistical principles for clinical trials ICH Harmonised Tripartite Guideline. 1995. *http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/statistical-principles-for-clinical-trials.html*.
26. International Conference on Harmonisation Expert Working Group. Structure and content of clinical study reports: E3. ICH harmonised tripartite guideline, 1995.
27. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. The American Statistician, 2016, 70(2): 129-133.
28. Stephens' dectection and evaluation of adverse drug reactions: Principles and practice. 6th ed. 2012, Chichester, West Sussex, UK: John Wiley & Sons, Ltd.

29. CIOMS Working Group X. Evidence synthesis and meta-analysis for drug safety: Report of the CIOMS working group X. 2016, Geneva, Switzerland: Council for International Organizations of Medical Sciences (CIOMS). 176.
30. Wisniewski AFZ, Bate A, Bousquet C, Brueckner A, Candore G, Juhlin K, Macia-Martinez MA, Manlik K, Quarcoo N, Seabroke S, Slattery J, Southworth H, Thakrar B, Tregunno P, Van Holle L, Kayser M, Norén GN. Good signal detection practices: Evidence from IMI PROTECT. Drug Safety, 2016, 39(6): 469-490.
31. Mehrotra DV, Heyse JF. Use of the false discovery rate for evaluating clinical safety data. Statistical Methods in Medical Research, 2004, 13: 227-238.
32. Berry SM, Berry DA. Accounting for multiplicities in assessing drug safety: A three-level hierarchical mixture model. Biometrics, 2004, 60: 418-426.
33. Ahlers C, Brueckner A, Mallick A, Opitz N, Pinkston V, Tran B, Scott J, Southworth H, Tran B, Van Holle L, Wallis N. Statistical signal detection in clinical trial data. in PROTECT Symposium. 2015. London, UK.
34. Schechtman E. Odds ratio, relative risk, absolute risk reduction, and the number needed to treat--which of these should we use? Value in Health, 2002, 5(5): 431-6.
35. Duke SP, Bancken F, Crowe B, Soukup M, Botsis T, Forshee R. Seeing is believing: Good graphic design principles for medical research. Statistics in Medicine, 2015.
36. Amit O, Heiberger R, Lane P. Graphical approaches to the analysis of safety data from clinical trials. Pharmaceutical Statistics, 2008, 7: 20-35.
37. Quade D, Lachenbruch PA, Whaley FS, McClish DK, Haley RW. Effects of misclassifications on statistical inferences in epidemiology. American Journal of Epidemiology, 1980: 503-515.
38. Copeland KT, Checkoway H, McMichael A, Holbrook RH. Bias due to misclassification in the estimation of relative risk. American Journal of Epiedmiology, 1977, 105(5): 488-495.
39. CIOMS Working Groups III and V. Guidelines for preparing core clinical-safety information on drugs. 1999, Geneva: Council for International Organizations of Medical Sciences.
40. Nilsson ME, Koke SC. Defining treatment-emergent adverse events with the Medical Dictionary for Regulatory Activities. Drug Information Journal, 2001, 35: 1289-1299.
41. Crowe B, Wang WV, Nilsson ME. Advances in techniques for combining data from multiple clinical trials, in Advances in collating and using trial data, M. Resche-Rigon and S. Chevret, Editors. 2014, Future Science. 20-37.
42. Crowe B, Xia HA, Nilsson ME, Shahin S, Wang WV, Jiang Q. The program safety analysis plan: An implementation guide. in Quantitative evaluation of safety in drug development: Design, analysis, and reporting, Q. Jiang and H.A. Xia, Editors. 2015, Chapman & Hall: London. 55-68.
43. Chuang-Stein C, Xia HA. The practice of pre-marketing safety assessment in drug development. Journal of Biopharmaceutical Statistics, 2013, 23(1): 3-25.
44. Mallick A. From AE to ADR: Medical aspects. 2012. Copenhagen, Denmark: 24th Annual DIA EuroMeeting.
45. U.S. Food and Drug Administration. Safety assessment for IND safety reporting: Guidance for industry. 2015. *http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm477584.pdf*.
46. Wittes J, Crowe B, Chuang-Stein C, Guettner A, Hall D, Jiang Q, Odenheimer D, Xia HA, Kramer J. The FDA's final rule on expedited safety reporting: Statistical considerations. Statistics in Biopharmaceutical Research, 2015, 7(3): 174-190.
47. Panel on Handling Missing Data in Clinical Trials and Committee on National Statistics, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, National Research Council. The prevention and treatment of missing data in clinical trials. 2010, Washington (DC): National Academies Press. 162.
48. European Medicines Agency. Guideline on missing data in confirmatory clinical trials. 2010. *http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf*.
49. Kraemer HC. Events per person-time (incidence rate): a misleading statistic? Statistics in Medicine, 2009, 28(6): 1028-1039.
50. Cook RJ, Lawless JF. The statistical analysis of recurrent events. 2007, New York, NY: Springer.
51. Bender R, Beckmann L, Lange S. Biometrical issues in the analysis of adverse events within the benefit assessment of drugs. Pharmaceutical Statistics, 2016.
52. Southworth H, O'Connell M. Data mining and statistically guided clinical review of adverse event data in clinical trials. Journal of Biopharmaceutical Statistics, 2009, 5: 803-817.
53. Firth D. Bias reduction of maximum likelihood estimates. Biometrika, 1993, 80(1): 27-38.
54. Fu H, Price KL, Nilsson ME, Ruberg SJ. Identifying potential adverse events dose-response relationships via Bayesian indirect and mixed treatment comparison models. Journal of Biopharmaceutical Statistics, 2013, 23(1): 26-42.

17. Appendix A: Tables and Figures Considered, but not Part of Recommendations

Table 17.1.
Summary of Treatment-Emergent Adverse Events by Preferred Term in Descending Frequency
Safety Population

Preferred Term	T1 (N=XX) n (%)	T2 (N=XX) n (%)	PL (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 treatment-emergent 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][a]	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][b]	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.

[a]Risk difference confidence interval from [insert method].
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX(Placebo) [For applicable PTs].

Table 17.2.
Summary of Treatment-Emergent Adverse Events Considered Related by the Investigator by Preferred Term in Descending Frequency
Safety Population

Preferred Term	T1 (N=XX) n (%)	T2 (N=XX) n (%)	PL (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 treatment-emergent adverse event considered related by the investigator	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][a]	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][b]	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.

[a]Risk difference confidence interval from [insert method].
[a]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [For applicable PTs].
[b]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX(Placebo) [For applicable PTs].

Table 17.3.
Summary of Exposure Adjusted Exposure-Adjusted Incidence Rates by Preferred Term in Descending Frequency within System Organ Class
Safety Population

Preferred Term	T1 (N=XX) n (%) / e [r]	T2 (N=XX) n (%) / e [r]	PL (N=XX) n (%) / e [r]	T1 & T2 (N=XX) n (%) / e [r]	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 treatment-emergent adverse event	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[System Organ Class #1]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]			
[Preferred Term #1]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[Preferred Term #2][a]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
[System Organ Class #2]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]			
[Preferred Term #1][b]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [xx.x]	xx (xx.x) / xx.x [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. Percentages are based on N
e = Sum across all patients, the total time to first event or total exposure if no event (years).
r = Exposure-adjusted patient rate per XXX patient years (n/e*XXX).

Patients may be counted in more than one row.

[a]Risk difference confidence interval from [insert method].
[b]Denominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo).
[c]Denominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo).

Table 17.4.
Treatment-Emergent Adverse Events by Sex

Production/Test Data – Production/Test Mode
Treatment-Emergent Adverse Events by Sex
Preferred Term by Decreasing Frequency within System Organ Class
All Evaluable Patients
Study, Phase

System Organ Class Preferred Term	Subgroup	Comparator			Treatment			Total			OR*a	P value*a	P value*b
		N	n	%	N	n	%	N	n	%	Between Treatments Within Subgroup	Between Treatments Within Subgroup	Trt* Subgroup
Patients with >= 1 TEAE	Female	xx	xx	(xx.x)	xx	xx	(xx.x)	xx	xx	(xx.x)	xx.xx	.xxx	.xxx
	Male	xx	xx	(xx.x)	xx	xx	(xx.x)	xx	xx	(xx.x)	xx.xx	.xxx	
System Organ Class 1	Female	xx	xx	(xx.x)	xx	xx	(xx.x)	xx	xx	(xx.x)	xx.xx	.xxx	.xxx
	Male	xx	xx	(xx.x)	xx	xx	(xx.x)	xx	xx	(xx.x)	xx.xx	.xxx	
Preferred Term 1-1	Female	xx	xx	(xx.x)	xx	xx	(xx.x)	xx	xx	(xx.x)	xx.xx	.xxx	.xxx
	Male	xx	xx	(xx.x)	xx	xx	(xx.x)	xx	xx	(xx.x)	xx.xx	.xxx	
Preferred Term 1-2	Female	xx	xx	(xx.x)	xx	xx	(xx.x)	xx	xx	(xx.x)	xx.xx	.xxx	.xxx
	Male	xx	xx	(xx.x)	xx	xx	(xx.x)	xx	xx	(xx.x)	xx.xx	.xxx	
System Organ Class 2	Female	xx	xx	(xx.x)	xx	xx	(xx.x)	xx	xx	(xx.x)	xx.xx	.xxx	.xxx
	Male	xx	xx	(xx.x)	xx	xx	(xx.x)	xx	xx	(xx.x)	xx.xx	.xxx	
Preferred Term n-n	Female	xx	xx	(xx.x)	xx	xx	(xx.x)	xx	xx	(xx.x)	xx.xx	.xxx	.xxx
	Male	xx	xx	(xx.x)	xx	xx	(xx.x)	xx	xx	(xx.x)	xx.xx	.xxx	

Abbreviations: N = number of evaluable patients; n = number of patients with treatment-emergent adverse event; OR = Mantel-Haenszel odds ratio; TEAE = treatment-emergent adverse event; Trt = Treatment;(additional if needed).
*a: Within group odds ratio from logistic regression model. Treatment is numerator, Comparator is denominator.
*b - P values are from logistic regression with model including study, subgroup, treatment and treatment*subgroup.
MedDRA Version X.X

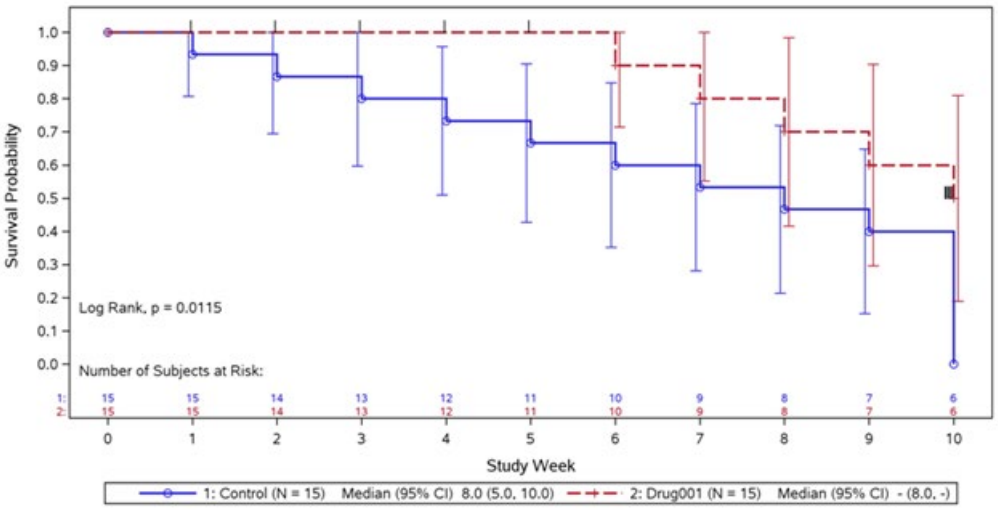
Table 17.5.
Summary of Subgroup Analysis for Possibly Related Grade 3 or 4 or 5 CTCAE Toxicities

Production/Test Data – Production/Test Mode
Summary and Analysis of Subgroup Analysis for
Possibly Related Grade 3 or 4 or 5 CTCAE Toxicities
Title (optional; add more if needed)
<Insert population (for example, Safety Population (N = xxx))>
<Insert study ID(s) or description of database utilized (for example, F3Z-MC-IOOX, F3Z-MC-IOOY,
or "All Acute Placebo-controlled Studies" or "Active-controlled Database")>
Study Phase or phases (if needed)

Grade 3 or 4 toxicity	Subgroup	Comparator		Treatment		Total		OR*a Within Subgroup	P value*b	
		N (%)	n (%)	N (%)	n (%)	N (%)	n (%)		Treatment* Subgroup	Within Subgroup
Patients with>=1 Grade 3/4/5 CTCAE	Group 1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x	.xxx	.xxx
	Group 2	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x		
CTCAE term 1	Group 1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x	.xxx	.xxx
	Group 2	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x		
CTCAE term 2	Group 1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x	.xxx	.xxx
	Group 2	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x		
CTCAE term n	Group 1	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x	.xxx	.xxx
	Group 2	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx (xx.x)	xx.x		

Abbreviations: CTCAE = common terminology criteria for adverse events; N = total number of patients; n = number of patients in the specified category with a given CTCAE; OR = odds ratio; add more as needed (alphabetically).
CTCAE Version X.X
*a - Within subgroup odds ratio from logistic regression model. Treatment is numerator, Comparator is denominator
*b - P values are from logistic regression with model include study, subgroup, treatment and treatment*subgroup.
Program Location: Program name
Output Location: Output name
Data Set Location: Data set name
Instructions:
Some columns/rows may be deleted or added as needed.
Subgroup parameters as needed. For studies that are for registration,

Figure 17.1.
Kaplan-Meier Estimates of Time to Event



Censor indicated by vertical bar |. Error bar represents a 95% pointwise confidence interval for the survivor function.

Figure 17.1.
Kaplan-Meier Estimates of Time to Event

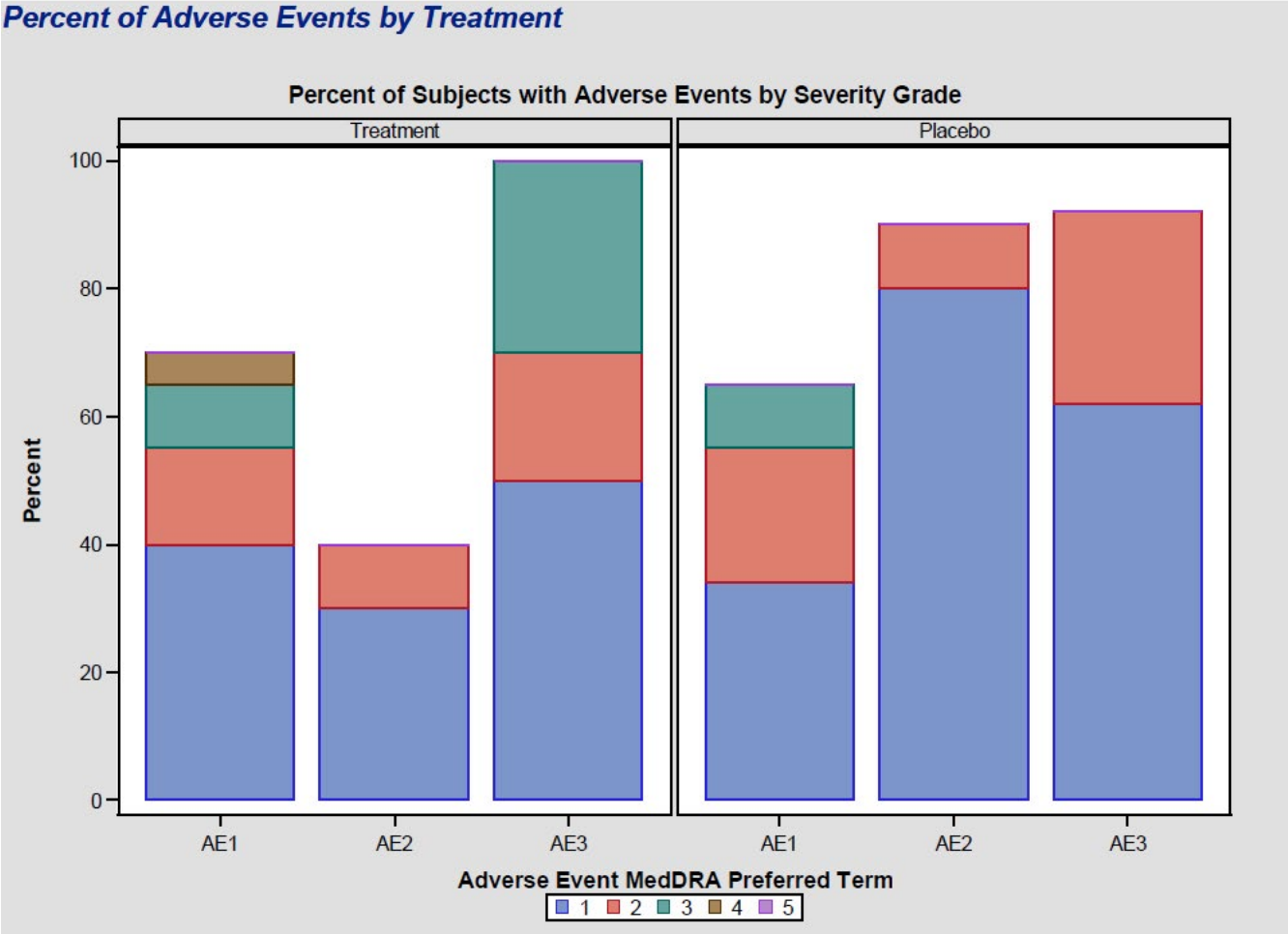


Figure 17.3.
Treatment-Emergent Adverse Events by Treatment Arm

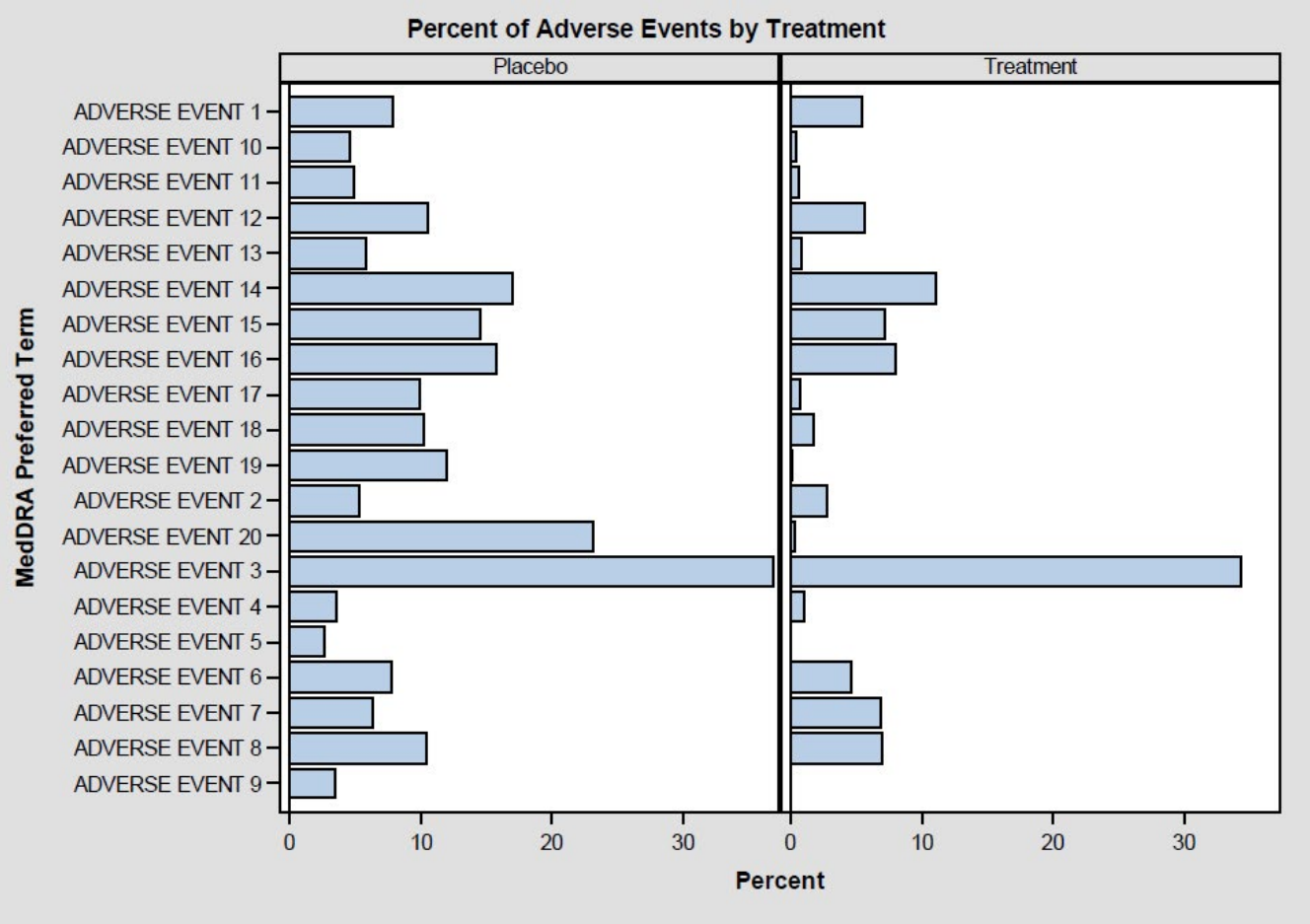
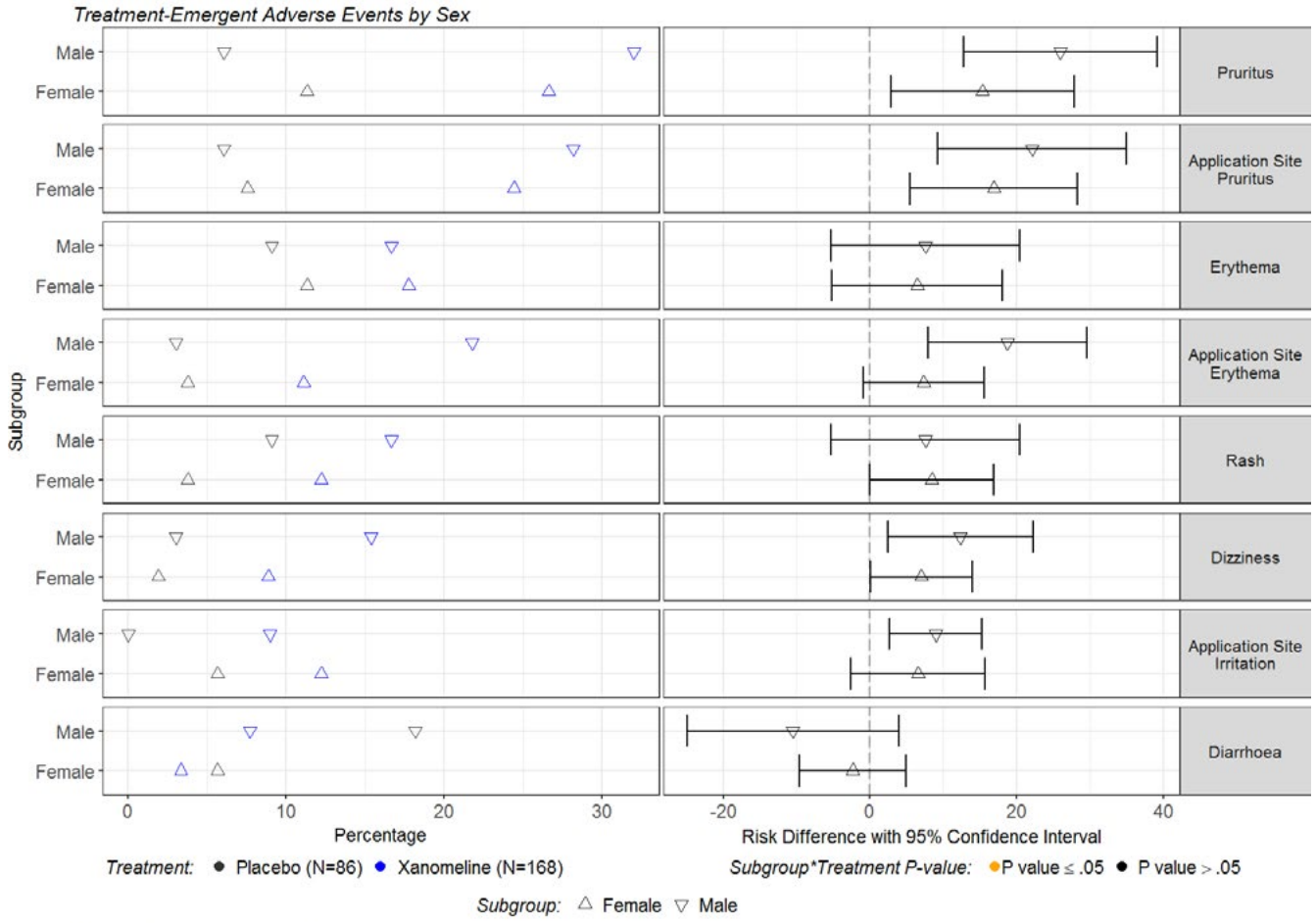
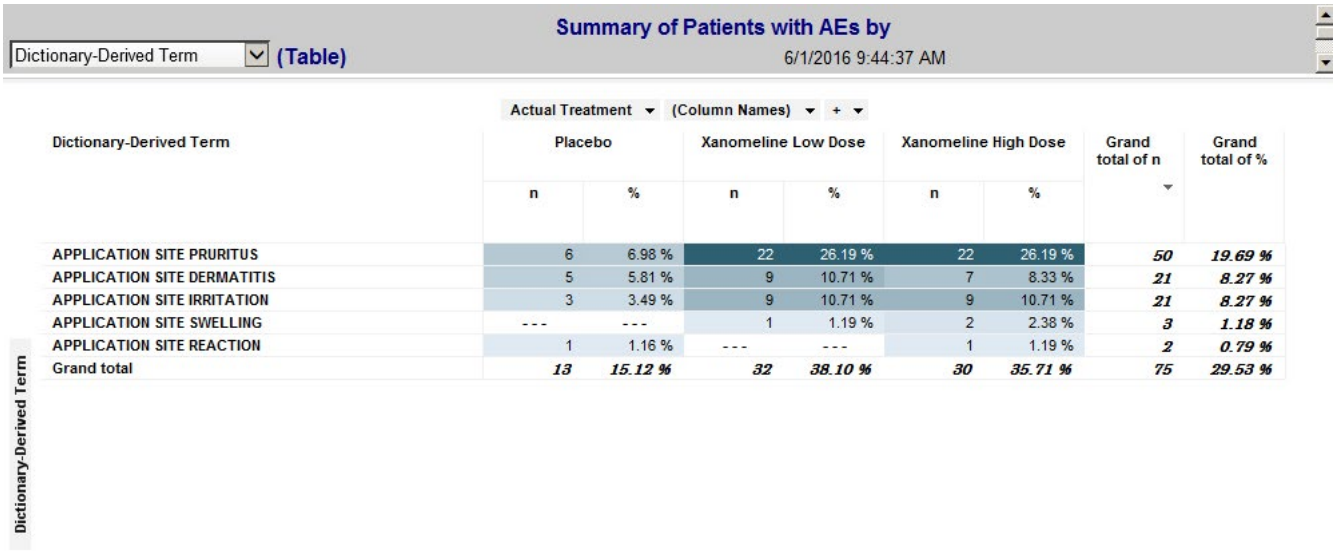


Figure 17.4.
Common Treatment-Emergent Adverse Events by Gender



Adverse Events are based on 10% in either treatment group.
P value is from the test of subgroup*treatment interaction in the logistic regression model.
The model includes treatment, subgroup and treatment*subgroup as independent variables.

18. Appendix B: Interactive Tool Snapshots



Summary of Patients with AEs by

Dictionary-Derived Term (Table)

6/1/2016 9:44:37 AM

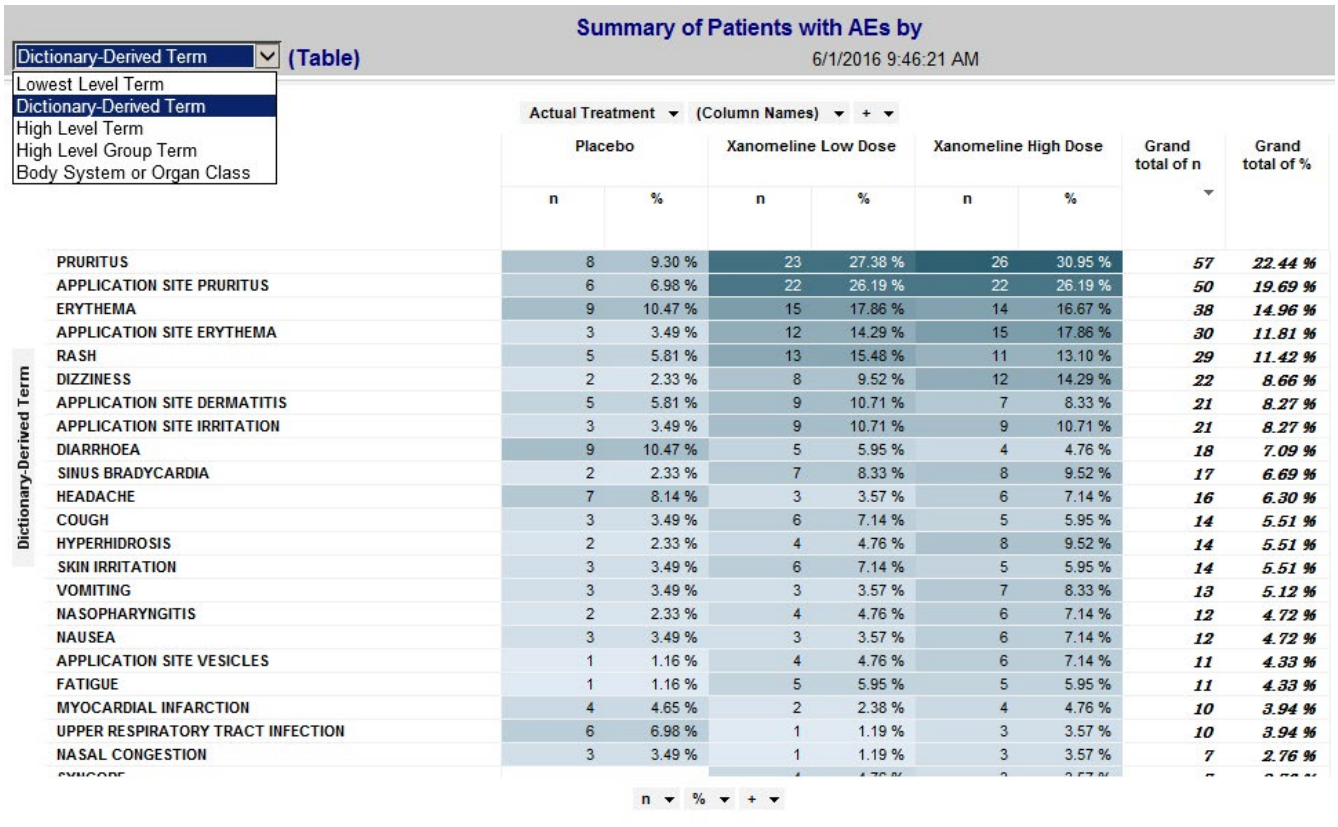
Actual Treatment (Column Names) +

Dictionary-Derived Term	Placebo		Xanomeline Low Dose		Xanomeline High Dose		Grand total of n	Grand total of %
	n	%	n	%	n	%		
APPLICATION SITE PRURITUS	6	6.98 %	22	26.19 %	22	26.19 %	50	19.69 %
APPLICATION SITE DERMATITIS	5	5.81 %	9	10.71 %	7	8.33 %	21	8.27 %
APPLICATION SITE IRRITATION	3	3.49 %	9	10.71 %	9	10.71 %	21	8.27 %
APPLICATION SITE SWELLING	---	---	1	1.19 %	2	2.38 %	3	1.18 %
APPLICATION SITE REACTION	1	1.16 %	---	---	1	1.19 %	2	0.79 %
Grand total	13	15.12 %	32	38.10 %	30	35.71 %	75	29.53 %

Dictionary-Derived Term

Example of an interactive display to group similar AEs:

- Pick which level to review the AEs (LLT, PT, HLT, HLGT, SOC).
- Choose the AEs to group.
- A display with each individual AE and with a Grand Total is created.



Summary of Patients with AEs by

Dictionary-Derived Term (Table)

6/1/2016 9:46:21 AM

Lowest Level Term
Dictionary-Derived Term
High Level Term
High Level Group Term
Body System or Organ Class

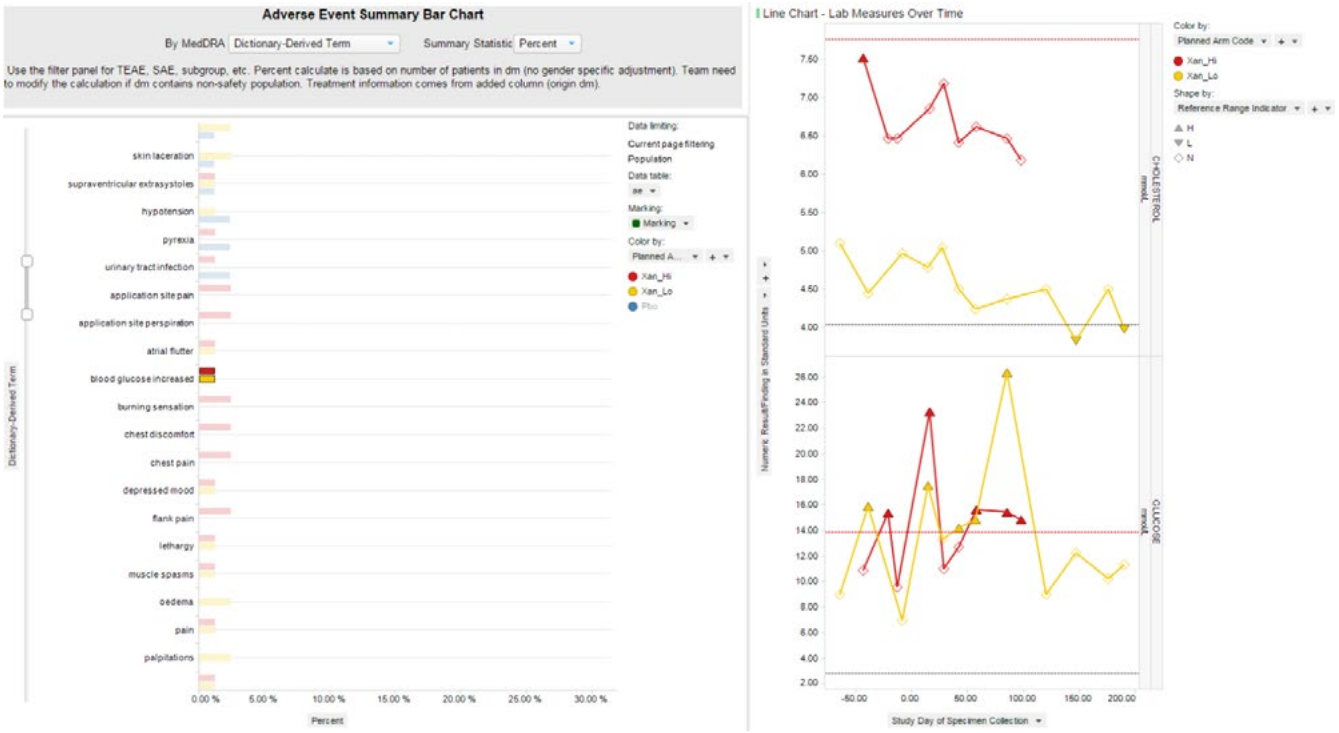
Actual Treatment (Column Names) +

Dictionary-Derived Term	Placebo		Xanomeline Low Dose		Xanomeline High Dose		Grand total of n	Grand total of %
	n	%	n	%	n	%		
PRURITUS	8	9.30 %	23	27.38 %	26	30.95 %	57	22.44 %
APPLICATION SITE PRURITUS	6	6.98 %	22	26.19 %	22	26.19 %	50	19.69 %
ERYTHEMA	9	10.47 %	15	17.86 %	14	16.67 %	38	14.96 %
APPLICATION SITE ERYTHEMA	3	3.49 %	12	14.29 %	15	17.86 %	30	11.81 %
RASH	5	5.81 %	13	15.48 %	11	13.10 %	29	11.42 %
DIZZINESS	2	2.33 %	8	9.52 %	12	14.29 %	22	8.66 %
APPLICATION SITE DERMATITIS	5	5.81 %	9	10.71 %	7	8.33 %	21	8.27 %
APPLICATION SITE IRRITATION	3	3.49 %	9	10.71 %	9	10.71 %	21	8.27 %
DIARRHOEA	9	10.47 %	5	5.95 %	4	4.76 %	18	7.09 %
SINUS BRADYCARDIA	2	2.33 %	7	8.33 %	8	9.52 %	17	6.69 %
HEADACHE	7	8.14 %	3	3.57 %	6	7.14 %	16	6.30 %
COUGH	3	3.49 %	6	7.14 %	5	5.95 %	14	5.51 %
HYPERHIDROSIS	2	2.33 %	4	4.76 %	8	9.52 %	14	5.51 %
SKIN IRRITATION	3	3.49 %	6	7.14 %	5	5.95 %	14	5.51 %
VOMITING	3	3.49 %	3	3.57 %	7	8.33 %	13	5.12 %
NASOPHARYNGITIS	2	2.33 %	4	4.76 %	6	7.14 %	12	4.72 %
NAUSEA	3	3.49 %	3	3.57 %	6	7.14 %	12	4.72 %
APPLICATION SITE VESICLES	1	1.16 %	4	4.76 %	6	7.14 %	11	4.33 %
FATIGUE	1	1.16 %	5	5.95 %	5	5.95 %	11	4.33 %
MYOCARDIAL INFARCTION	4	4.65 %	2	2.38 %	4	4.76 %	10	3.94 %
UPPER RESPIRATORY TRACT INFECTION	6	6.98 %	1	1.19 %	3	3.57 %	10	3.94 %
NASAL CONGESTION	3	3.49 %	1	1.19 %	3	3.57 %	7	2.76 %

Dictionary-Derived Term

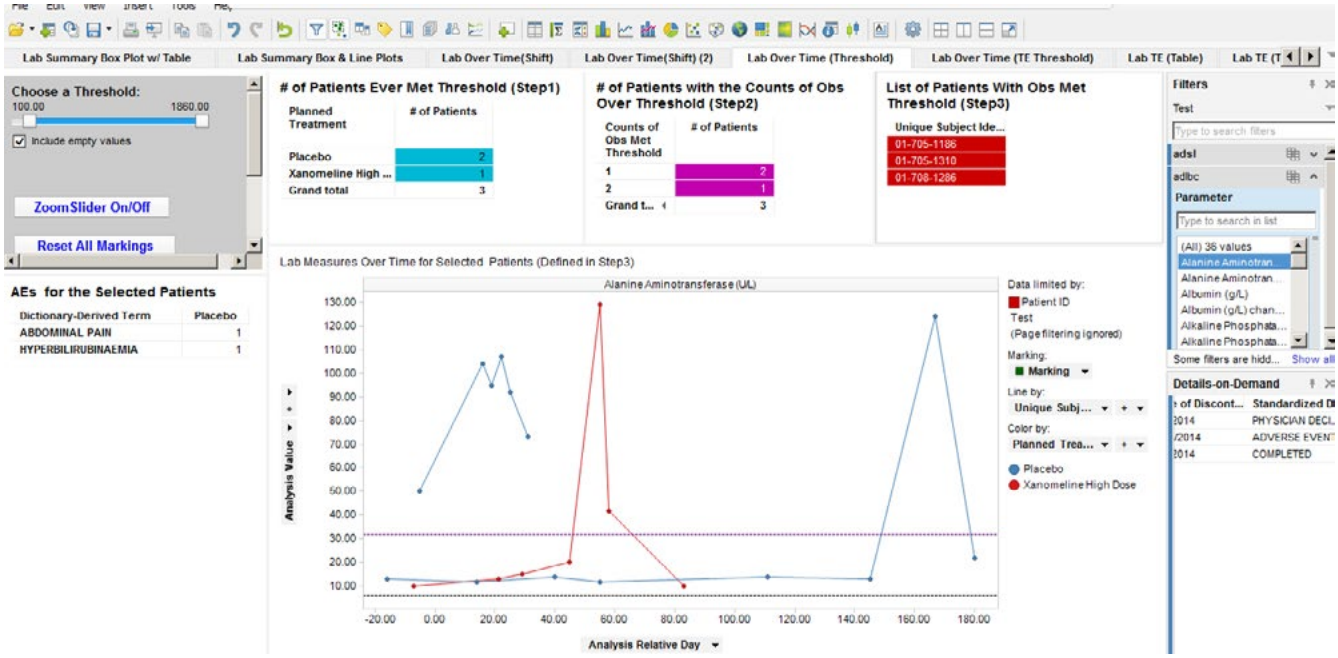
n % +

Example of an interactive display of AEs. Shows the drop-down menu on how the user can choose the level to review.



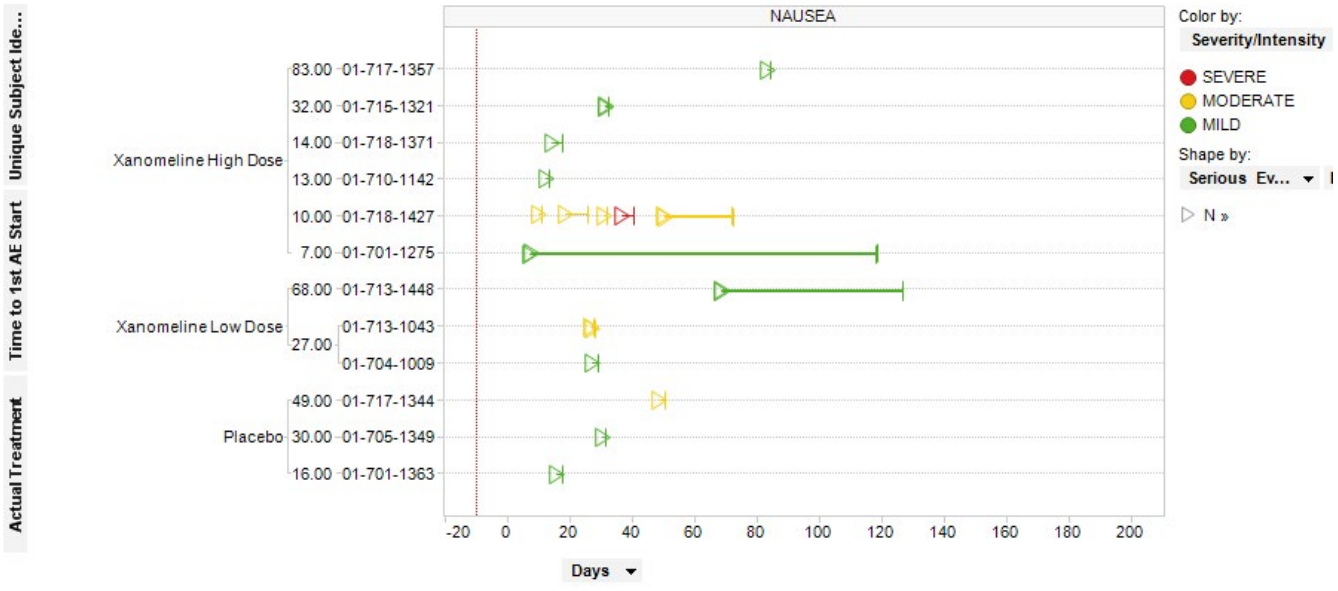
Example of an interactive display for AEs with relationship to labs:

- Pick which level to review the AEs (LLT, PT, HLT, HLGT, SOC).
- Pick a summary statistic.
- A graph of the AE data is displayed.
- If the reviewer would like to see lab data for a particular AE, highlight the bars for the AE. Pick the lab parameters of interest and the data will be added to the display.

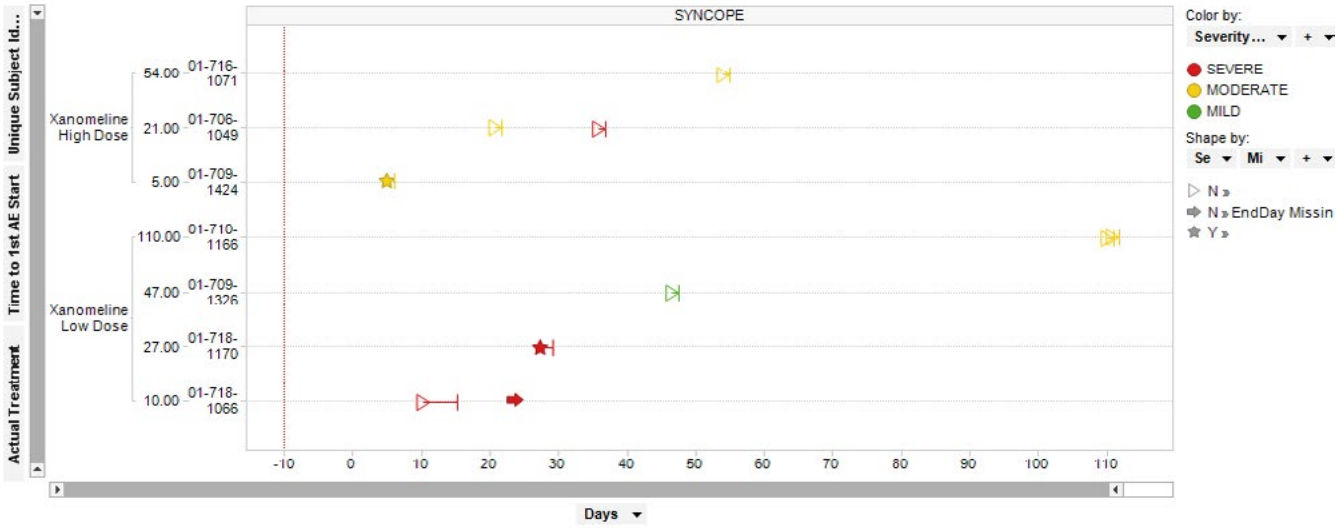


Example of an interactive display for patients having a lab exceeding a threshold:

- Pick a lab parameter (right side).
- Choose a threshold (first box).
- Highlight the count of patients that have at least one lab value meeting the threshold (2nd box), or the count of patients who met a particular count over the threshold.
- Hit Enter.
- Highlight the patients.
- A line of the lab data is displayed. The AEs for the patients are also shown in the lower left panel.



Example of an interactive display of AEs over time. The user chooses the event. This provides an assessment on whether the event tends to start early versus late (or random) and how long the event lasts. The top plot is an example of an event that changes in severity. The bottom plot is an example that shows that a different symbol can be used when the event is serious.


SYNCOPE
Unique Subject Id...
Time to 1st AE Start
Actual Treatment
Xanomeline High Dose
Xanomeline Low Dose
54.00
21.00
5.00
110.00
47.00
27.00
10.00
01-716-1071
01-706-1049
01-709-1424
01-710-1166
01-709-1326
01-718-1170
01-718-1066
-10
0
10
20
30
40
50
60
70
80
90
100
110
Days
Color by:
Severity...
SEVERE
MODERATE
MILD
Shape by:
Se
Mi
N »
N » EndDay Missin
Y »