

Safety Analytics Workshop

PhUSE CSS

09 June 2019

Speaker Introduction – Mary Nilsson



- MS degree in statistics from Iowa State University in 1989
- Employed at Lilly since 1989
- Approximately 8 submissions
- Safety Analytics group since 2012 (consults with compound teams on safety analysis planning)
- Co-leads the PhUSE Standard Analyses and Code Sharing Working Group
- Leads the PhUSE Analysis and Display white papers project team

Acknowledgements

- Co-authors and reviewers of the PhUSE white papers
- Analysis and Display white papers project team
- Nhi Beasley (FDA Liaison)
- Wei Wang, Rebeka Revis
- Lori VanMeter, Jerry Xu, Swarna Reddy, Brenda Crowe

Pre-reads

- The following white papers from the PhUSE Deliverables Catalog are listed as suggested pre-reads:
 - [AE white paper](#)
 - [Labs/Vitals/ECGs Central Tendency white paper](#)
 - [Labs/Vitals/ECGs Outlier/Shift white paper](#)
 - [Demographics/Disposition/Medications white paper](#)
- Additional useful resource:
 - [FDA 2010 Clinical Review Template](#) (Attachment B, starting on page B-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.

Agenda

- Background
- Definitions and Purpose
- Adverse Events
- Labs
- Medications
- Summary

This workshop will be recorded
and publicly posted

Outline for Adverse Events, Labs, Medications

- Discuss recommended tables and figures
 - Individual study tables and figures
 - Expand on rationale
 - Expand on interpretation of tables and figures
 - Share common pitfalls
 - Cover frequently asked questions



Background for Having a Workshop

- There's a gap in safety analytics knowledge across all functions (including statistics)
- Educating during internal team meetings and during regulatory interactions is difficult – one agenda of many, usually during time pressure

Myths

- Analytical planning for safety is easy
- Interpreting safety results is easy



Analysis and Display White Paper (ADW) Project

- Lead: Mary Nilsson
FDA Liaison: Nhi Beasley
- Description: This project includes the development of white papers that provide recommended Tables, Figures, and Listings for clinical trial study reports and submission documents. The intent is to begin the process of developing industry standards with respect to analysis and reporting for measurements that are common across clinical trials and across therapeutic areas

ADW Project Team Goals

- Improve expertise in safety analytics across multiple disciplines involved with planning, interpreting, and reporting safety analyses
 - Promote good analytical practices, avoid poor practices
- Define a set of recommended tables, figures, listings and associated analyses
 - Ensure reviewers receive clinically relevant and meaningful analyses for benefit-risk assessment
 - Harmonization leads to efficiencies in both creation and review of analyses
- Facilitate cross-functional engagement in analytical planning

ADW – Final Deliverables

- All final white papers from the project team are in the Working Group [Deliverables Catalog](#)
 - White Papers section
- Adverse Events (2017)
 - Vitals Signs, Labs, ECGs
 - Central Tendency (2013)
 - Outliers/Shifts (2015)
 - Demographics, Disposition, Medications (2014, 2018)
 - Non-compartmental PK (2014)
 - QT/QTc Studies (2016)

Definitions to Facilitate Discussion

- **Signal detection** – Analyses and displays used to determine which events or changes in labs/vitals/ECGs require further evaluation for adverse drug reaction (ADR) consideration
- **Signal clarification** – Additional analyses and displays used to facilitate the assessment of whether an event or change in labs/vitals/ECGs should be identified as an ADR
- **Signal characterization** - Additional summaries used to characterize events or changes in labs/vitals/ECGs once they are determined to be an ADR
- **Signal communication** - Additional summaries used for documents that communicate ADRs (eg, regional labelling)

Definitions to Facilitate Discussion

- **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).
- **Interactive display** – A table, figure, or listing that allows point-and-click technology to zoom, scroll, hover over a data point to see more information, or control information displayed (eg, choose among odds ratio, risk ratio, risk difference)

Convention to Facilitate Discussion

- The FDA/EMA/PMDA says”, or “the FDA/EMA/PMDA thinks”
 - Let’s reserve this language for when the position is documented in a regulatory guidance document
- “Our clinical reviewers asked”, or “some clinical reviewers think”
 - Let’s use this language when the position hasn’t been established in a regulatory guidance document

Common Pitfall: Assuming an opinion held by a regulatory reviewer is a commonly held opinion across the agency

Adverse Drug Reaction (ADR) Definition

- 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

ADR Determination

Common Pitfall: Thinking the list is determined based on some rule (eg, at least 1% and greater than placebo)

- Factors used to determine the list of ADRs
 - Strength of evidence for an imbalance between study drug and placebo (via p-values or confidence interval, or sometimes the “eyeball” test)
 - Magnitude of the observed effect (via odds ratio, risk ratio, or risk difference)
 - Observed dose relationship
 - Biologic plausibility
 - Clinical relevance of any individual case (e.g., any available dechallenge/rechallenge information)
 - Severity of the event
 - Consistency of findings across studies
 - Consistency of findings across similar events
 - Consistency of findings across similar compounds

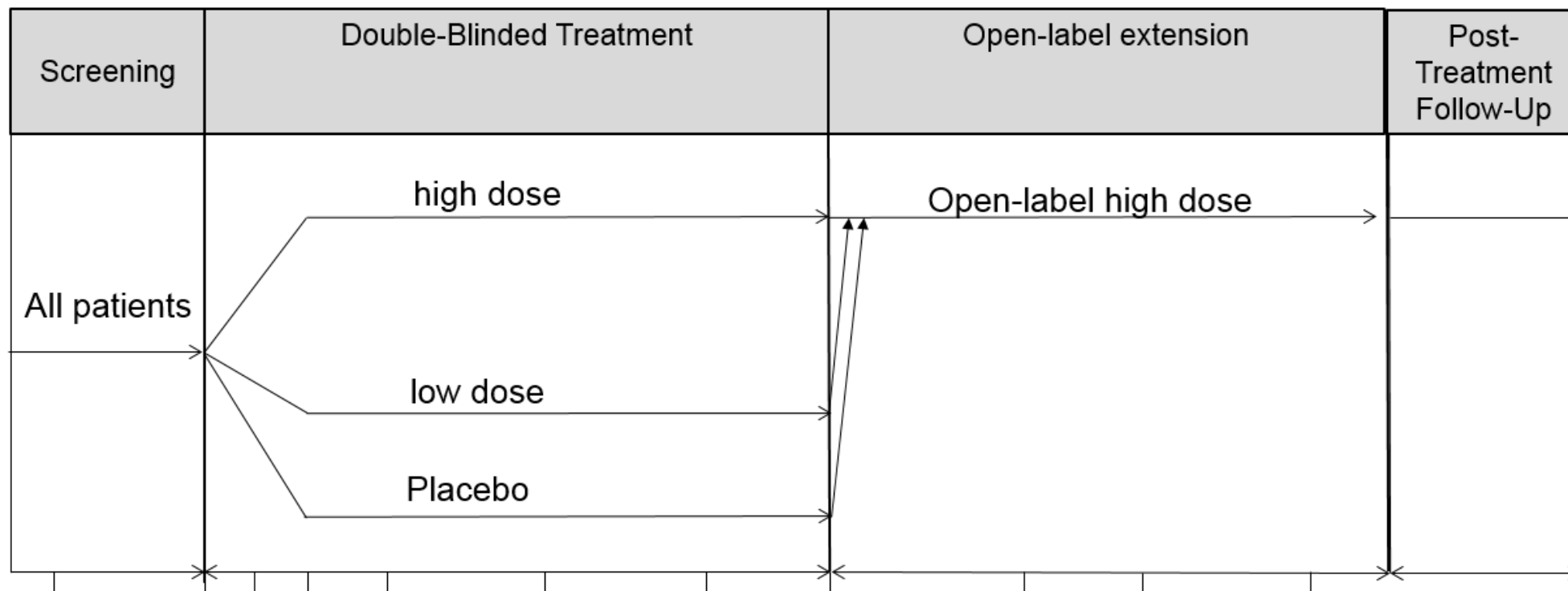
Purpose of Initial Set of Safety Analyses

- Signal detection
 - Safety topics of interest have separate analysis plans, usually signal detection and clarification

Common Pitfalls:

- A-priori creating analyses for all 4 types (detection, clarification, characterization, communication)
- Applying what is needed for a safety topic of interest (eg, clarification) to all safety
- Applying what is needed for a label (eg, communication) to all safety

Study Design Used for Discussion



Adverse Events – The List (from AE white paper)

Study Drug vs Placebo

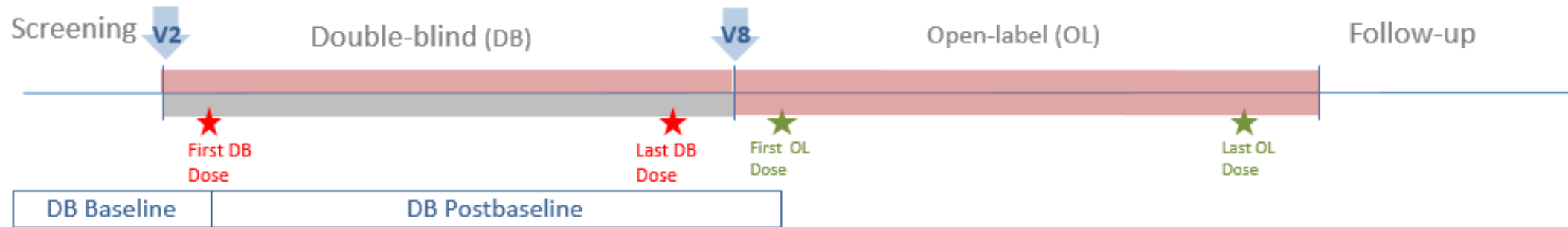
- ◆ Overview of AEs
- ◆ TEAEs sorted by PT within SOC
- ◆ TEAEs sorted by PT
- ◆ Common TEAEs (eg, $\geq 1\%$)
 - usually a table but prefer a figure
- ◆ TEAEs by maximum severity
- ◆ AEs leading to study treatment discontinuation (PT **or** PT within SOC)
- ◆ SAEs (PT **or** PT within SOC)

During Study Drug and During Study Drug+After

- ◆ Overview of AEs
- ◆ TEAEs sorted (PT **or** PT within SOC)
- ◆ AEs leading to study treatment discontinuation (PT **or** PT within SOC)
- ◆ SAEs (PT **or** PT within SOC)

Percentages

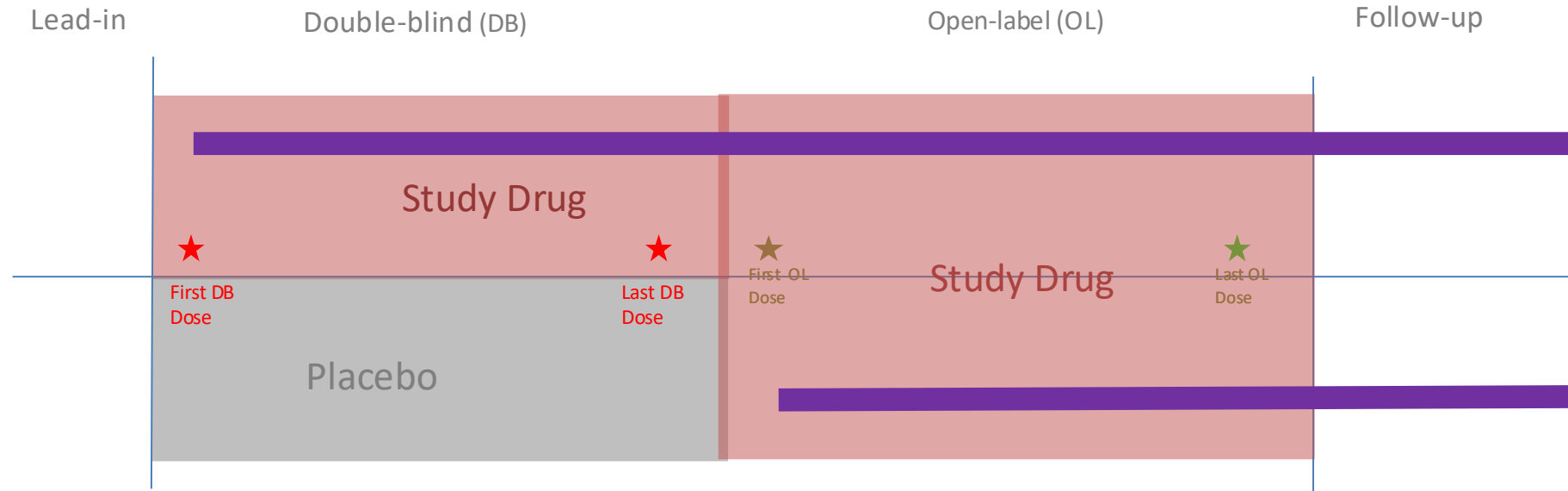
Adverse Events – Time Periods



Baseline: All time before the first DB Dose

Postbaseline: Time from first DB dose to right before OL dose or the summary visit in DB period, whichever comes first

Adverse Events – Time Periods



The above illustrates the cohorts and what part are included in the 'During Study Drug Treatment and After' analysis. **Both cohorts will be pooled together in the analysis.**

Adverse Events – Treatment Period Shell

T1 = Low dose
T2 = High dose
PL=placebo

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.

^aRisk difference confidence interval from [insert method].

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

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

Adverse Events – Maximum Severity Shell

T1 = Low dose
T2 = High dose
PL=placebo

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:

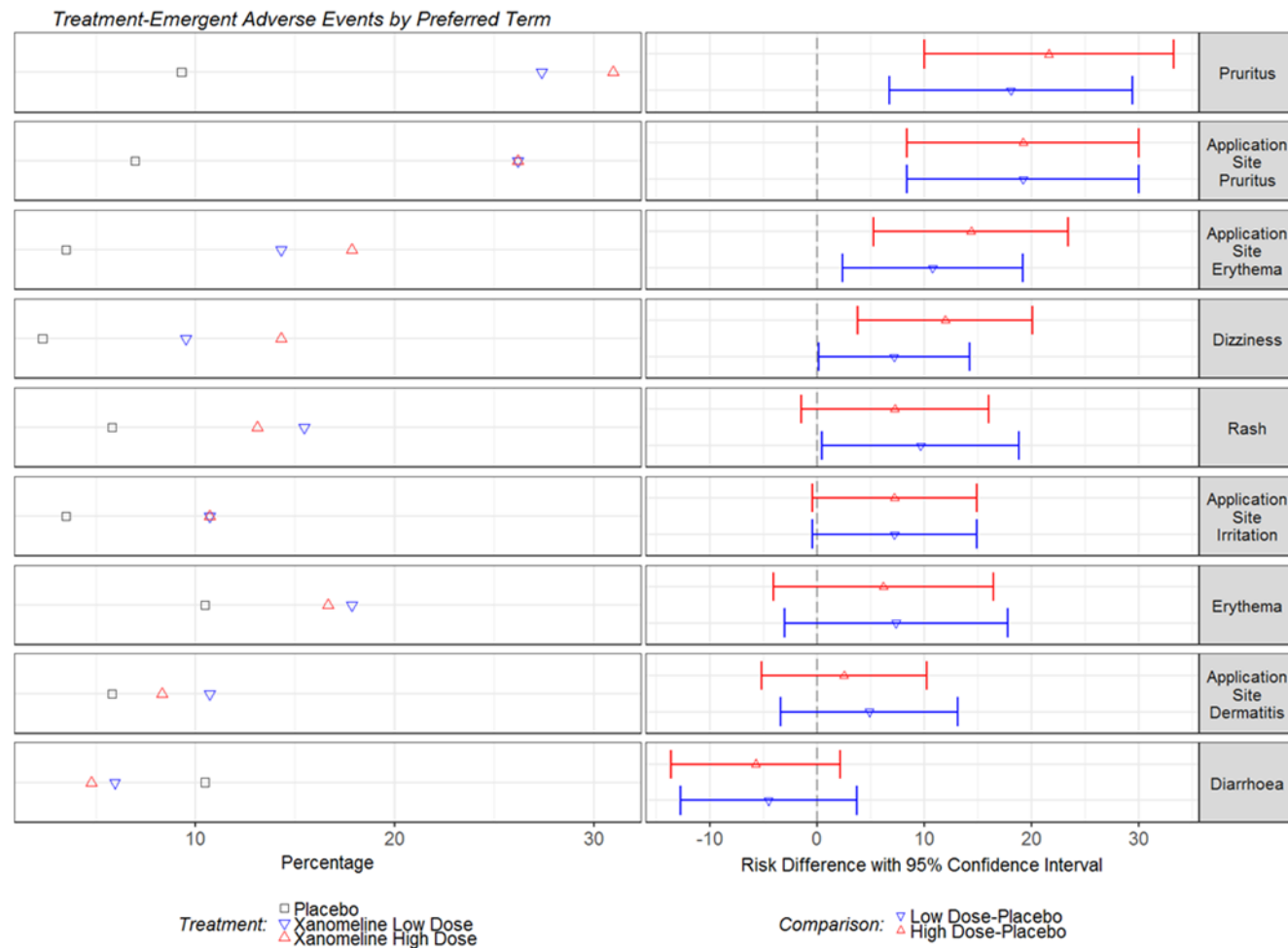
N=number of patients; n=number of patients with at least one row event. Percentages are calculated relative to the treatment group N. Patients may be counted in more than one row.

^aRisk difference confidence interval from [insert method].

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

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

Adverse Events – Treatment Period Shell



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

Adverse Events – During Study Drug and During Study Drug+After

Table x.x. Treatment-Emergent Adverse Events

System Organ Class Preferred Term	During Study Drug (N=XX) n (%)	During Study Drug+After (N=XX) n (%)
Number of subjects reporting at least one treatment-emergent adverse event	xx (xx.x)	xx (xx.x)
[System Organ Class #1]	xx (xx.x)	xx (xx.x)
[Preferred Term #1]	xx (xx.x)	xx (xx.x)
[Preferred Term #2] ^a	xx (xx.x)	xx (xx.x)
[System Organ Class #2]	xx (xx.x)	xx (xx.x)
[Preferred Term #1] ^b	xx (xx.x)	xx (xx.x)
...		

N=number of subjects; n=number of subjects with at least one row event.
Subjects may be counted in more than one row.
^aDenominator adjusted because sex-specific event for males: N=XX [For applicable PTs].
^bDenominator adjusted because sex-specific event for females: N=XX [For applicable PTs].

Adverse Events – Rationale

- Study Drug vs Placebo
 - Provides strength of evidence for an imbalance with placebo and magnitude of effect (part of ADR determination)
- During Study Drug and During Study Drug+After
 - During SD: Provides a way to see all events – looking for the more rare events, and late onset events that might require case reviews
 - During SD+After: Same as During SD except takes the conservative approach by assuming any event after treatment could be due to treatment

Adverse Events – Rationale

Table	Rationale
Overview of AEs	Provides a nice summary for CSR purposes
TEAEs sorted by decreasing frequency across PTs within SOC	Used to look for events with strong evidence for an imbalance, high magnitude of effect, or noteworthy based on clinical judgment
TEAEs sorted by decreasing frequency across PTs	Same as above, but offers a different way of sorting
Common TEAEs sorted by decreasing frequency across PTs	We don't really have a scientific rationale, but templates call for this
TEAEs by maximum severity	Including all events can obscure a meaningful signal – looking at severe events only provides another chance for finding a signal
AEs leading to study treatment discontinuation (one sort)	Same as TEAEs, max severity
SAEs (one sort)	Same as TEAEs, max severity

Adverse Events - Rationale

- TEAEs by maximum severity
 - See the FDA Clinical Review Template for a discussion on the potential for missing signals when looking at all severities for events with high background rates

Common Pitfall: Using the maximum severity table for signal characterization only and not review the full table in the context of signal detection, and not include a signal-detection related statement in CSRs

Adverse Events - FAQs

- How is treatment emergence defined?
 - See Section 10.2 of the PhUSE AE white paper for a brief discussion and references
 - Recommending a specific definition is out-of-scope for the AE white paper and this workshop
 - Planned for a future white paper

Common Pitfall: Assuming your way of defining treatment emergence is predominant practice

Adverse Events - FAQs

- Why do we count the number of patients with events instead of number of events?
 - Simpler, works fine in most cases
 - Variability in event collection (eg, each headache versus “intermittent headaches”) complicates interpretability of number of events
 - Counting events is likely warranted for some safety topics of interest (eg, hypoglycemia, infections), but generally not warranted as part of signal detection

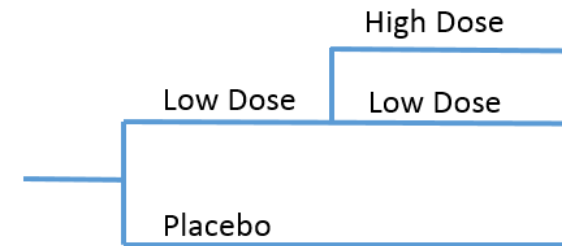
Adverse Events - FAQs

- Why is the list for During Study Drug and During Study Drug+After shorter?
 - Common TEAEs – The During SD group is more about the more rare AEs; need to look at all events; sub-setting on common is not value-added for the During SD group
 - Max Severity – Included in the placebo-controlled group primarily to compare severe PTs between SD and placebo as part of signal detection
 - Time to study treatment discontinuation due to AE – Not as useful without a comparator

Adverse Events - FAQs

- Why isn't dose included in the During Study Drug summary?
 - Dose relatedness is best assessed from the initial treatment period in which randomization is preserved
 - Due to channeling bias, summaries by dose including data across periods often lead to misleading interpretations about dose relatedness
 - In many designs, randomization is no longer preserved, and biases are usually substantial enough such that summaries are illogical (based on a lot of shared learning)

Common pitfall: Thinking it's value-added to create summaries by dose using data across multiple periods of a study



Adverse Events - FAQs

- Why have confidence intervals or p-values?
 - Having a measure assessing strength of evidence for an imbalance is helpful and is part of ADR determination
 - For safety, think of p-values as an assessment of strength of evidence for an imbalance, using a scale from 0 to 1
 - Not used for hypothesis testing
 - Confidence intervals provide a means for assessing strength of evidence for an imbalance, but not as easy to use for “ranking” strength and/or facilitating consistent thinking for events with similar strength

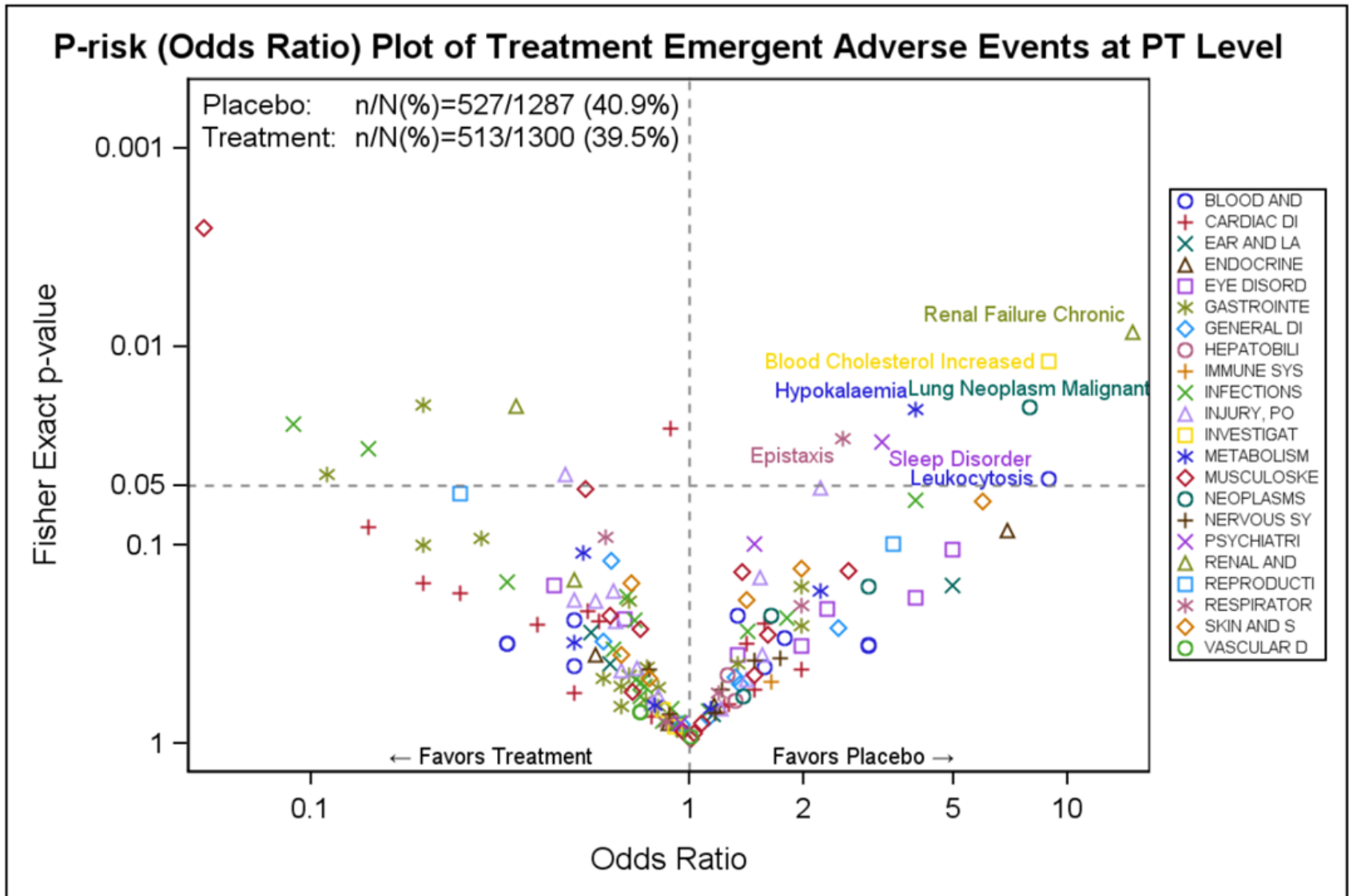
Common pitfall: Focusing on events with p-values less than a specified value (eg, 0.05) or confidence intervals that don't include 0 (for risk differences) or 1 (for risk ratio, odds ratio)

Adverse Events - FAQs

- Strength of evidence for an imbalance and magnitude of effect are not the same
 - Example: Odds ratios are the same, p-values are different

Event	Study Drug	Placebo	OR	P-value
PT 1	700/1500 (46.7%)	550/1500 (36.7%)	1.5	<.001
PT 2	41/1500 (2.7%)	27/1500 (1.8%)	1.5	.110

Volcano Plot –
Shows both
strength of
evidence for an
imbalance and
magnitude of
effect



Adverse Events - FAQs

- Why does postbaseline end at the study treatment disposition visit (for those who discontinue)?
 - Conceptually, it shouldn't
 - Estimating the percent of patients with an event during the treatment period would suggest all data during the treatment period should be included
 - However, in practice, it seems patients are allowed to take other medications once study treatment is discontinued
 - For signal detection purposes (study drug vs placebo), the bias created by the addition of events potentially due to other medications is usually considered worse than the bias created by excluding events occurring during the time between study treatment discontinuation through the end of the treatment period.

Adverse Events - FAQs

- Why didn't you add 30 days (or some other number based on the half-life of the compound) to the placebo-controlled analysis?
 - Usually interested in estimating the percent of patients with an event after x weeks of treatment (where x is the length of the treatment period)
 - If 30 days were added, you're estimating the percent of patients with an event after x weeks of treatment plus 30 days off treatment
 - In the example design, if 30 days were added to both study drug and placebo, study drug events would attributed to placebo. You could exclude events occurring after switching to another drug, but that introduces bias by doing something to one arm and not the other.

Common Pitfall: Assuming a convention for adding days is predominant practice

Adverse Events - FAQs

Common Pitfall: For safety analyses, choosing the most conservative approach across various choices

- Conservative from a total count perspective may lead to being less conservative from a treatment difference perspective
- To assess treatment differences, recommend making analytical decisions that lead to the least amount of bias

Adverse Events - FAQs

- Why don't you have analyses for the OL period by itself, and the follow-up period by itself?
 - If data are presented by period, it makes it more difficult to find the more rare events that warrant further scrutiny
 - Unless the compound has withdrawal effects, a summary of the follow-up period by itself isn't usually associated with an estimate of interest
 - There's a better way to assess withdrawal effects
 - There's a better way to characterize events after stopping treatment

Common Pitfall: Plan for all safety analyses by period for all periods

Adverse Events - FAQs

- Why didn't you include exposure-adjusted incidence rates?
 - For controlled periods, we don't usually have differential observation time, so exposure-adjustment isn't usually warranted
 - For During Study Drug, exposure-adjusted incidence rates might be useful, but we recommend the additional summaries for integrated only

Common Pitfall: Thinking that exposure-adjustment is a complete solution for taking care of differences in observation time

Common Pitfall: Under-appreciation that exposure-adjusted incidence rates can perform poorly for events that mainly occur early

Common Pitfall: Creating exposure-adjusted incidence rates that includes off-treatment time

Adverse Events - FAQs

- Why do some compound teams use subjective severity levels (mild/moderate/severe) when CTC objective levels can be used?
 - See the PhUSE AE white paper for a discussion
 - Developed for oncology since oncology patients experience many adverse events and such severity granularity is considered useful
 - For most indications, subjective severity levels seem sufficient (and considered less burdensome)
 - Knowing seriousness and knowing whether the event led to treatment discontinuation are additional indicators of severity

Adverse Events - FAQs

- Why didn't you include a summary of AEs considered related by the investigator?
 - Assessments by investigators are highly dependent on information available at the time of the clinical trial (eg, what is in the investigator's brochure)
 - Thus, it's generally not useful to use their assessment as part of signal detection
 - Could be useful in the review of individual cases for signal clarification

Adverse Events - Interpretation

- Study Drug vs Placebo comparisons
 - Confidence intervals and p-values provide an assessment on how strong the evidence is for an imbalance – NOT used to assess statistical significance
 - See [March 2019 special issue of The American Statistician](#)
 - Moving to a world beyond $p < 0.05$
 - *As “statistical significance” is used less, statistical thinking will be used more*

Common Pitfall: Statements such as “The percentage was higher in study drug vs placebo, but the confidence interval included 1”

Adverse Events - Interpretation

- During Study Drug and During Study Drug+After data
 - With additional exposure time, there is more opportunity to see more of the rare or late onset events that could require further scrutiny for ADR determination
 - Interpretation is primarily based on medical judgment, with knowledge of the population, knowledge of the design, length of study, and case reviews
 - During Study Drug+After interpreted as conservative count (attributes an event after study drug to study drug)

Common Pitfall: Comparing During Study Drug percentages or exposure-adjusted incidence rates with placebo-controlled percentages

Adverse Events - Interpretation

- Concept of primary
 - Unlike efficacy, defining a primary safety analysis set is not needed
 - Placebo-controlled data will be the most useful for the more common events
 - Longer-term data will be the most useful for late onset and the more rare events

Common Pitfall: Identifying a primary safety analysis set

Adverse Events - Interpretation

- Discuss events that are noteworthy based on clinical judgment, have strong evidence for an imbalance, or high magnitude of effect
 - The ADR determination factors can be considered for the discussion, however a lot of discussion is most appropriate for an integrated summary
 - For individual studies, only need conclusions for the study, not for the compound

Common Pitfall: Focus discussion on common events

Adverse Events - Interpretation

- Be aware of splitting
 - From FDA Clinical Review Template 2010: The reviewer should consider the following:
 - *Whether the terms are too narrow (splitting), resulting in an underestimation of the true incidence for a particular event or syndrome (e.g., somnolence, drowsiness, sedation, and sleepiness probably all refer to the same event)*
 - “Clustering” or “Consolidation” is recommended at time of integrated summaries

Adverse Events - Interpretation

- Be aware of lumping

- From FDA Clinical Review Template 2010: The reviewer should consider the following:

Whether the terms are too broad or over-inclusive (lumping), so that important events that should be examined separately are diluted by less important events (e.g., loss of consciousness and syncope subsumed under hypotensive events or hypotension)

- Review of event descriptions during trial level safety reviews is key

Common Pitfall: Lack of appreciating the importance of reviewing event descriptions for specificity, and sending site queries where needed

Questions?



Labs – The List (from labs white papers)

- Study Drug vs Placebo
 - Shifts to high/low (performing lab limits)
 - Scatterplots
 - Minimum values
 - Maximum values
 - Boxplots over time
 - Observed
 - Change (includes change to last)
 - Shift tables
 - Minimum values
 - Maximum values
 - During Study Drug and During Study Drug+After
 - Shifts to high/low
- Not all displays would need to go in the CSR
 - Some displays could be created via an interactive tool

Labs – Common Pitfalls

Creating summaries for all labs in the database (as opposed to just those intended for group-level summaries)

Creating summaries for labs collected in a reflex manner

Creating summaries for labs in which we only receive abnormalities

Lesson learned: If summaries are warranted for labs in which we only receive abnormalities (cell morphology, urine microscopic exam), special data derivations are required (eg, to impute normal records)

Labs – Treatment Period Shell

T1 = Low dose
T2 = High dose
PL=placebo

Table x.x. Shifts to High/Low for Laboratory Measurements

	T1 (N=XX) n/N1 (%)	T2 (N=XX) n/N1 (%)	PL (N=XX) n/N1 (%)	T1 & T2 (N=XX) n/N1 (%)	Risk Difference ^a T1-PL (95% CI)	Risk Difference ^a T2-PL (95% CI)	Risk Difference ^a T1&T2-PL (95% CI)
Lab #1							
High	x/xx (xx.x)	x/xx (xx.x)	x/xx (xx.x)	x/xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
Low	x/xx (xx.x)	x/xx (xx.x)	x/xx (xx.x)	x/xx (xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)	xx.x (xx.x, xx.x)
...							

N=number of subjects; N1=Number at risk; n=number of subjects with at least one row event. Percentages are calculated relative to the treatment group N1.

Patients may be counted in more than one row.

^aRisk difference confidence interval from [insert method].

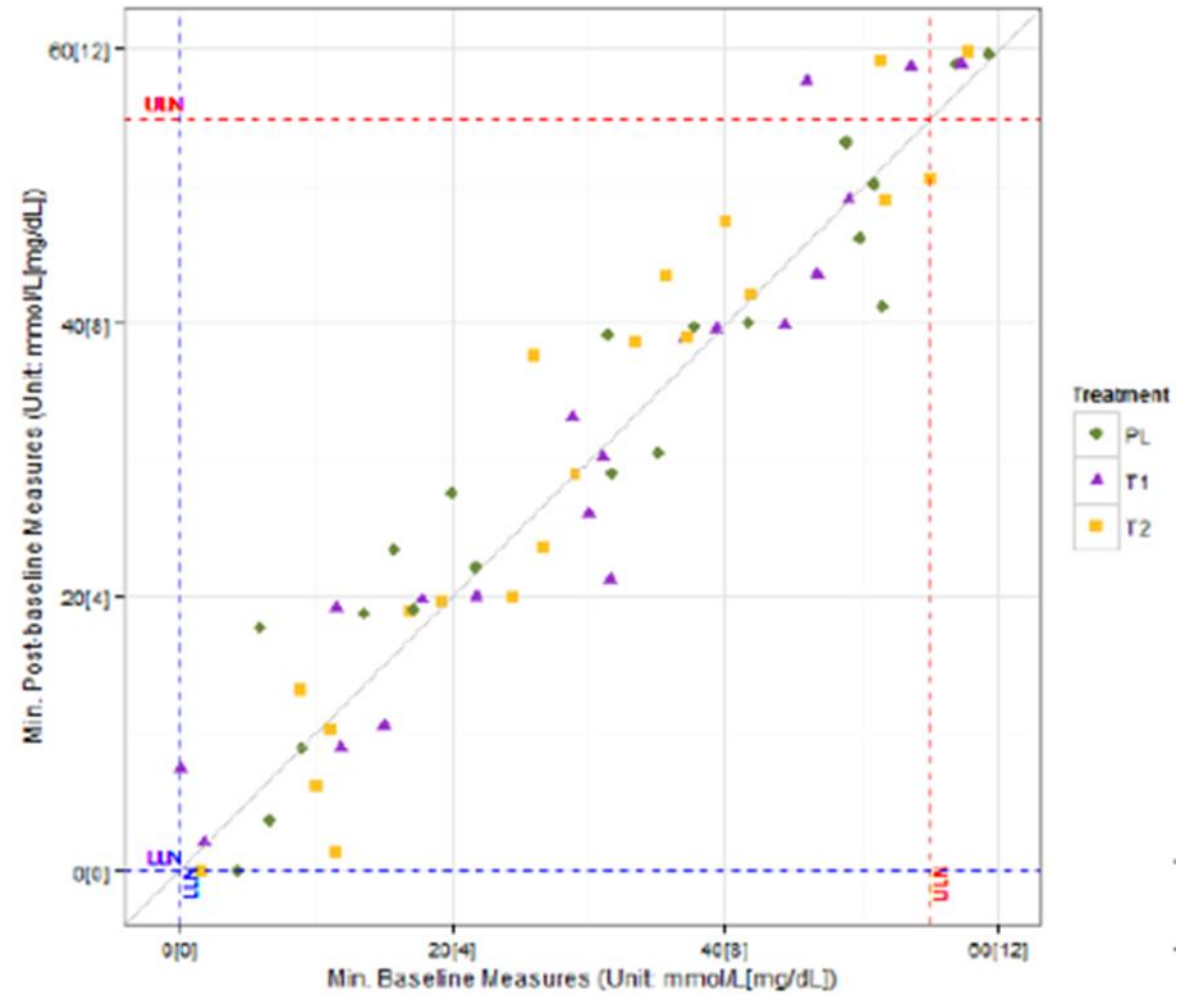
^bDenominator adjusted because sex-specific event for males: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [applicable PTs].

^cDenominator adjusted because sex-specific event for females: N=XX (Treatment 1), N=XX (Treatment 2), N=XX (Placebo) [applicable PTs].

Labs – Treatment Period Shell - Scatterplot

Both SI and CN units
can be included in
legend

T1 = Low dose
T2 = High dose
PL=placebo



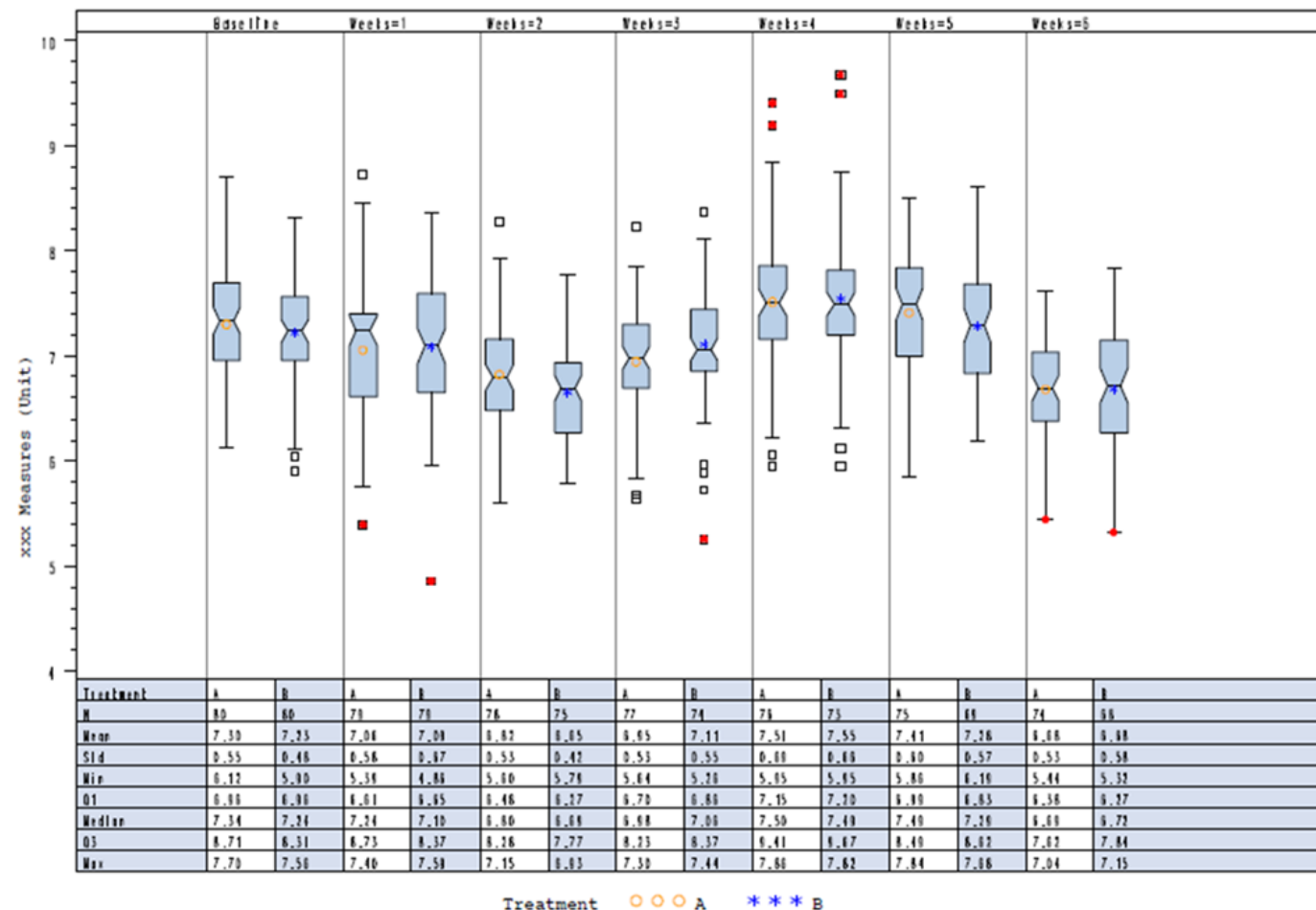
Labs – Treatment Period Shell - Boxplot

Both SI and CN units
can be included in
legend

Summary table –
SI units

Common pitfall: Forget
to keep only planned
measurements

Box Plot - xxx Measures Over Time (Weeks Since Randomized)/Treatment Emergent Change



Box plot type=schematic, the box shows median,interquartile range (IQR, edge of the box), min and max within 1.5 IQR below 25% and above 75% (ends of the whisker). Values outside the 1.5 IQR below 25% and above 75% are shown as outliers. Means plotted as different symbols by treatments. Red dots indicate out of normal reference range measures.

Labs – Treatment Period Shell – Shift Table

Scatterplot and Shift Table Summary of Absolute Lab Values - Lab Test 1
Maximum Baseline vs. Maximum Post-baseline

Treatment	Max. Baseline Result	Max. Post-baseline Result (Shift from Baseline)			
		Low n (%)	Normal n (%)	High n (%)	Missing n (%)
T1 (N = xxx)	Low	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
	Normal	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
	High	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
	Missing	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
T2 (N = xxx)	Low	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
	Normal	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
	High	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
	Missing	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
PL (N = xxx)	Low	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
	Normal	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
	High	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
	Missing	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)

N = number of subjects in the analysis population; n = number of subjects in each shift category; High and low are determined using the reference limits based on the individual subject's demographics.

Labs – During Study Drug and During Study Drug+After

Table x,x. Shifts to High/Low for Laboratory Measurements

System Organ Class Preferred Term	During Study Drug (N= <u>XX</u>) n/N1 (%)	During Study <u>Drug+After</u> (N= <u>XX</u>) n/N1 (%)
Lab #1	x/xx (<u>xx,x</u>)	x/xx (<u>xx,x</u>)
High	x/xx (<u>xx,x</u>)	x/xx (<u>xx,x</u>)
Low	x/xx (<u>xx,x</u>)	x/xx (<u>xx,x</u>)
...		

N=number of subjects; N1=Number at risk; n=number of subjects with at least one row event. Percentages are calculated relative to the treatment group N1.
Subjects may be counted in more than one row.

Labs - Rationale

- Part of ADR determination (same as adverse events)
- Need both central tendency and outlier/shift for controlled portion of the study
 - Outlier/shift most helpful, but could miss signals without central tendency
 - 2010 FDA Clinical Review Template – uric acid, electrolyte levels provided as examples

Labs - Rationale

Table/Figure	Rationale
Shift to high/low	Used to look for shifts with strong evidence for an imbalance, high magnitude of effect, or noteworthy based on clinical judgment
Scatterplot	Easy to see extreme values; can visually see patients contributing to TE high/low and how much they are shifting, can determine if there's a clustering just under a performing lab limit that would be worth exploration
Boxplot	Of the many options for displaying changes in labs in a central tendency manner, this seems to be a good choice as visually seeing changes over time is desirable, and can visually see extreme values
Shift table	Not considered key to signal detection, but currently recommended due to its popularity; may be removed from recommendations in the next version of the PhUSE white paper
Change to last observation	Not considered useful to signal detection, but currently part of predominant practice; may be removed from recommendations in the next version of the PhUSE white paper

Common pitfall: Considering change to last observation as part of signal detection

Labs - FAQs

- Why do we have to look at so many displays for lab signal detection?
 - Most summaries have a purpose
 - Some are recommended due to a perceived high risk of being asked as part of a regulatory request if not produced
 - Ongoing thought on how to focus/improve (version 2 of the white paper in progress)
 - Plans for lab collection and analysis can evolve across the lifecycle of a compound
 - See the [FDA 2016 late stage guidance](#)

Reminder - Not all displays would need to go in the CSR

Labs - FAQs

- For the treatment periods summaries, why does the scatterplot and shift table use max/min baseline instead of alternative choices?
 - Tempers the impact of within-patient variability
 - Using all measures prior to treatment (usually 2), instead of the last or average, tempers the impact of having more measures during treatment
 - Aligns with Lilly's treatment-emergent adverse event definition
 - Compares max severity during treatment with max severity during baseline/screening

Labs - FAQs

- Since efficacy data are usually analyzed using MMRM, why aren't you recommending MMRM for safety?
 - Primary efficacy analyses often use MMRM to estimate a mean difference at the end of the treatment period
 - Safety not interested in estimating this
 - If done, would use MMRM to look for group average differences over time, but alternatives seem more reasonable
 - Change to max/min seems to fit signal detection better (part of integrated recommendations), and allows the inclusion of unplanned measurements
 - Could be considered for labs of special interest (signal clarification)

Labs - FAQs

- Why do you exclude unplanned/unscheduled labs from the boxplots?
 - For any display over time, it's easier to see trends over time without the disruption of a few visits with very few patients
 - Inclusion of such data could create bias toward normality
 - See Section 10.3 of the PhUSE outliers/shifts white paper for further discussion

Labs - FAQs

- Why worry about CN units? Hasn't the research community aligned on SI units?
 - Collection and analysis would be so much easier if all physicians were used to the same units!
 - Unfortunately, that does not reflect reality
 - See a [2013 FDA letter](#) on the topic
 - Having different units creates a very complicated data flow from collection through analysis (eg, when to round, precision in rounding, units for limits, order of when to compare with limits)
 - We are constantly learning and adjusting standards
 - It's painful, but we're stuck with this pain

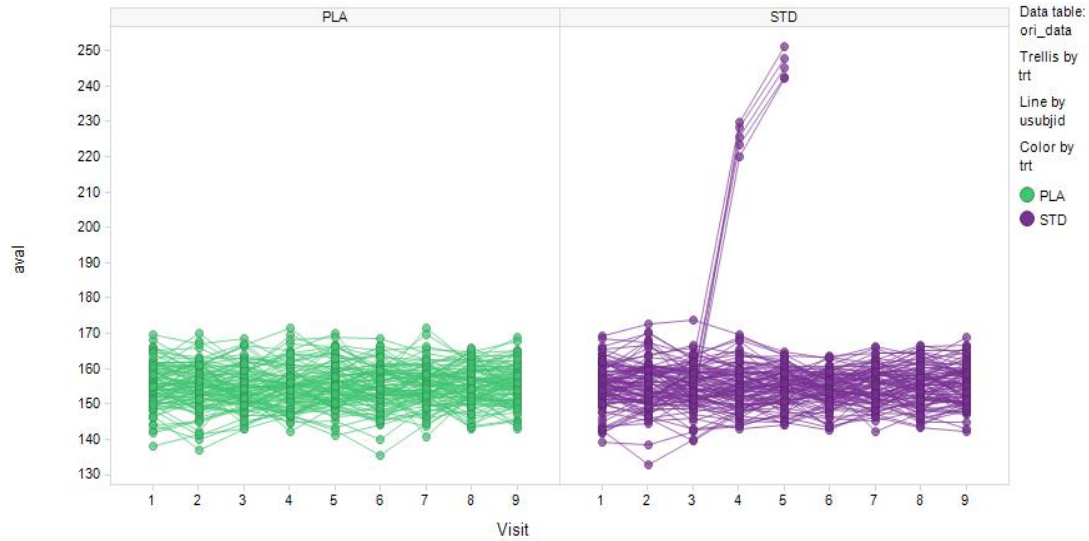
Labs - Interpretation

- Most topics discussed for AEs apply to labs
- Boxplots
 - When assessing clinical significance of a shift in a mean, more thought is needed than thinking about what that change might mean for an individual patient
 - Not reliable for transient versus persistent assessments
 - Spaghetti plots are useful for this

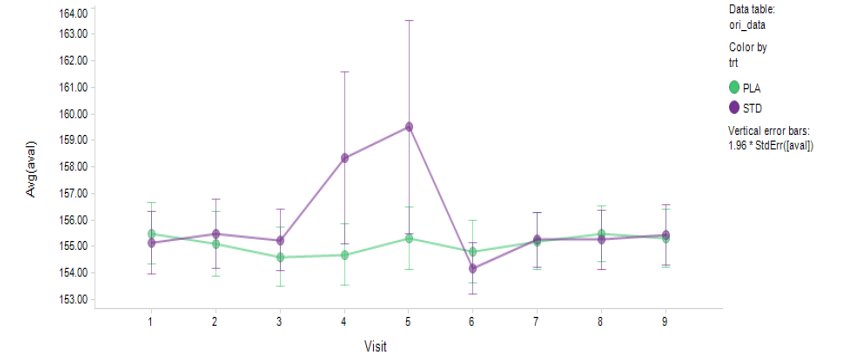
Common pitfall: Considering a result as not being clinically meaningful by reflecting on whether the change would be meaningful for an individual

Common pitfall: Using group means to assess persistent/transient patterns

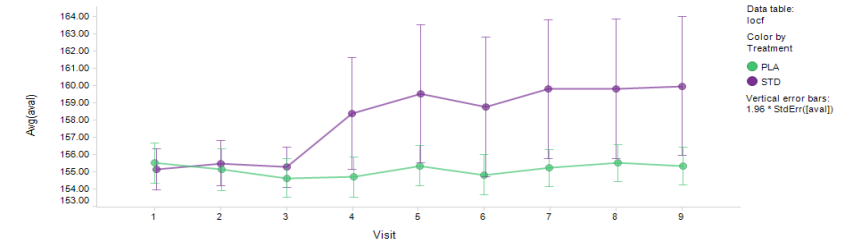
Labs - Interpretation



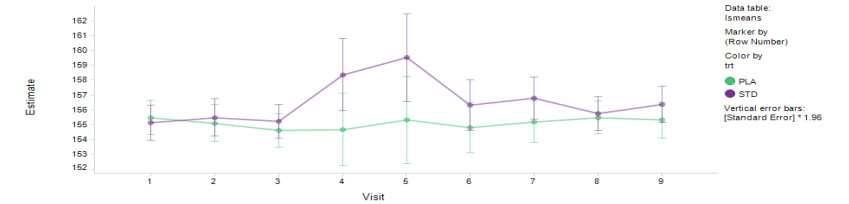
Observed



LOCF



MMRM



Questions?



Medications – The List (from demographics, disposition, medications white paper)

Study Drug vs Placebo

- ◆ Summary of prior medications – preferred names
- ◆ Summary of concomitant medications – preferred names

During Study Drug and During Study Drug+After

- ◆ None

Percentages

Medications – Rationale

Study Drug vs Placebo

- ◆ Understand the population under study more
- ◆ Assess whether there is an imbalance in medication use that would be important to know when reviewing safety data

During Study Drug and During Study Drug+After

- ◆ None

Medications - FAQs

- Since you don't recommend a summary of concomitant medications beyond the controlled treatment period, why collect medications during open-label or off-treatment periods?
 - Could be important for individual case reviews
 - Patient profiles include the time course of study drug and concomitant medications with the time course of adverse events, laboratory measurements, etc.

Common pitfall: Forgetting about the all-important patient profile. Just because data might not be used in a table, figure, listing in a study's List of Analyses doesn't mean it isn't important for individual case reviews.



Medications - FAQ

- Why do you recommend summaries using preferred names instead of ATC codes?
 - See the demographics, disposition, medications white paper for a thorough discussion
 - The same medication can be mapped to different ATC codes depending on reported intended use (unlike MedDRA that has primary mappings defined)
 - Not ideal for the intended purpose of the summary table

Disposition

- There's a recommended listing of subjects who discontinue due to physician decision, withdrawal by subject, withdrawal by parent/guardian, or other
 - See the demographics, disposition, medications white paper for a thorough discussion
 - Provides assurance that these subjects discontinued for reasons unrelated to adverse events or efficacy (by listing what is entered in a “specify” field)
 - Prevents the need for defining these subjects as “notable”

Integrated Summaries

- See white papers for recommendations
 - Many recommended displays are similar to those for individual studies, except with different statistical methods (eg, stratification by study)
 - Out-of-scope for this workshop
 - Upcoming workshop and/or additional educational videos are planned
- See educational video on study-size adjusted percentages
 - Go to phuse.eu, click on Working Groups, click on PhUSE references, go to the Standard Analyses and Code Sharing Section, click on “Topics for pooling data from multiple studies”

Common pitfall: Using improper methods when pooling data from multiple studies

Summary

- Safety planning and interpretation is not easy
 - Requires expertise not usually obtained prior to joining a team
 - Requires intermediate to advanced scientific thinking
 - Keep the purpose of the display in mind
 - Think about what you are estimating
 - Keep any comparisons unbiased as possible
 - Requires cross-functional discussion
- Improve your safety analytics expertise – read guidance/papers
- Get Engaged!
 - Review draft white papers
 - Version 2 of labs white paper is under development
 - Safety topics of interest white paper in development



Questions?





phuse.eu

**The premier community
for people working in the
biometrics area**

 @phusetwitta
 @phusesnappa
 /phusebook

 /phusetube
 /groups/887477/
 /phusepin

SACS Working Group Vision

- Vision: Leverage crowd-sourcing to improve the content and implementation of analyses for medical research, leading to better data interpretations and increased efficiency in the clinical development and review processes

SACS Working Group Goals

- Where gaps exist, develop recommendations for analyses and displays in areas that could benefit from crowd-sourcing
- Establish and maintain a publicly available repository for storing program code to be used as analytical tools for medical research
- Where gaps exist, develop code for recommended analyses and displays that could benefit from crowd-sourcing (to reside in the repository)
- Where gaps exist, develop guidance for good programming, quality, and validation practices

Current Projects

Analysis and Display White
Papers (ADW)
Lead: Mary Nilsson

Communication, Promotion,
Education (CPE)
Leads: Wendy Dobson, Jared Slain

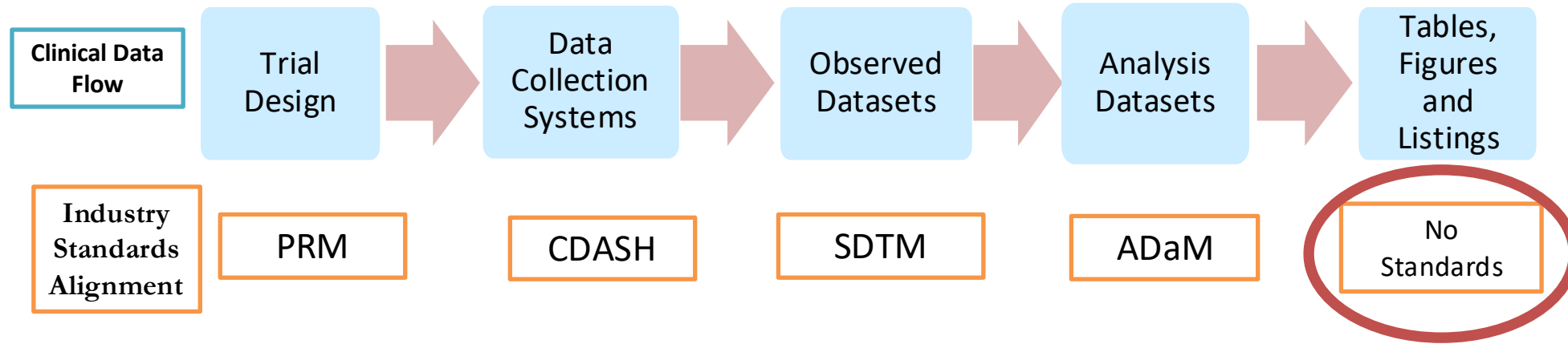
Macro Development
Lead: Ninan Luke

Code Sharing (Repository)
Leads: Hanming Tu, Bob
Friedman, Gustav Bernard

Best Practices for Quality Control and
Validation
Leads: Maria Dalton and Jane Marrer

Test Data Factory
Lead: Peter Schaefer

ADW Project Team Vision



PRM = Protocol Representation Model

CDASH = Clinical Data Acquisition Standards Harmonization

SDTM = Study Data Tabulation Model

ADaM = Analysis Data Model

ADW - Planned Deliverables

White Paper	Public Review Target
ADW style guide	NA
General output tips and considerations	Q3 2019
Safety topics of interest	Q3 2019
Labs	Q3 2019
Vitals	Q4 2019
Questionnaires	Q4 2019
Hepatotoxicity	Q4 2019
Listings in CSRs and submissions	Q1 2020
Treatment emergent definitions	Q1 2020
Interactive displays for clinical safety data	Q2 2020
ECGs	Q3 2020

ADR Evolution

- [1994 ICH E2A](#): In the pre-approval clinical experience with a new medicinal product or its new usages, particularly as the therapeutic dose(s) may not be established: all noxious and unintended responses to a medicinal product related to any dose should be considered adverse drug reactions. The phrase "responses to a medicinal products" means that a causal relationship between a medicinal product and an adverse event is at least a reasonable possibility, i.e., the relationship cannot be ruled out.
- 1995+ CIOMS III and V: Avoid including events, especially minor events, that have had no well-established relationship to therapy. As previously noted, the primary focus of the CCSI should be the description of **adverse reactions**. However, the Working Group felt that it could be useful to include a tabulation of the most frequently reported **adverse events** by drug compared with placebo.
- [2006 FDA ADR Guidance](#): For purposes of prescription drug labeling and this guidance, an adverse reaction is an undesirable effect, reasonably associated with the use of a drug, that may occur as part of the pharmacological action of the drug or may be unpredictable in its occurrence. This definition does not include all adverse events observed during use of a drug, only those for which there is some basis to believe there is a causal relationship between the drug and the occurrence of the adverse event.

ADR Evolution (continued)

- [2010 FDA Clinical Review Template](#): Adverse reaction refers to an undesirable effect, reasonably associated with the use of a drug, that may occur as part of the pharmacological action of the drug or may be unpredictable in its occurrence. This term does not include all adverse events observed during or after use of a drug, only those for which there is some basis to believe there is a causal relationship between the drug and the occurrence of the adverse event.
- [2012 FDA guidance on safety reporting requirements](#): An adverse reaction means any adverse event caused by a drug. Adverse reactions are a subset of all suspected adverse reactions where there is reason to conclude that the drug caused the event.
- [2017 PhUSE AE white paper](#): 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 Events - FAQs

- Why does the table with maximum severities not have confidence intervals for the mild and moderate rows?
 - Relates to the primary purpose of the table – checks that we don't miss a signal if we focus on the more severe events
 - Comparing percentages between treatments of exactly mild and exactly moderate are not generally clinically useful comparisons

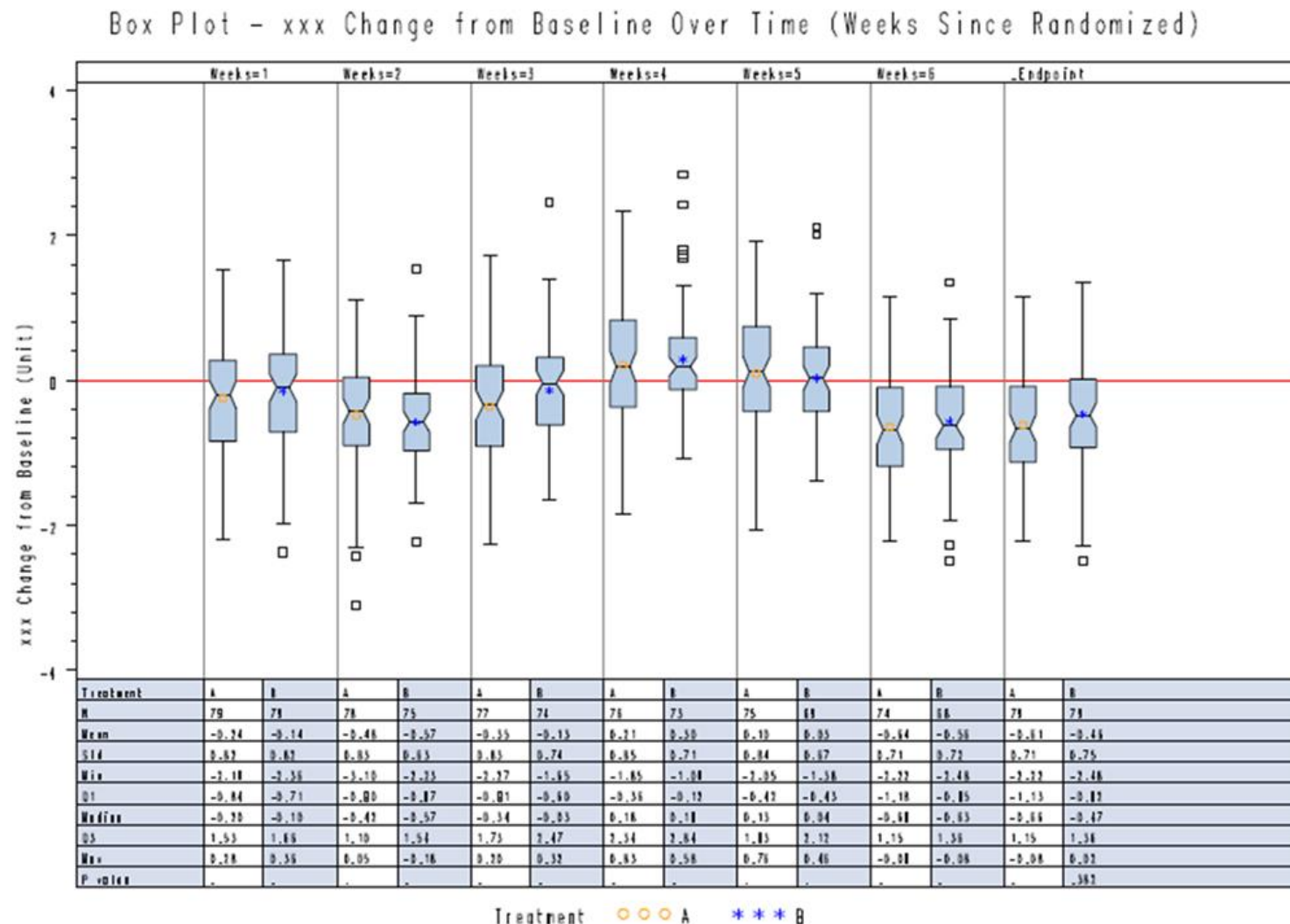
Adverse Events - FAQs

- Why do you recommend 2 ways of sorting for TEAEs, but not SAEs and AEs leading to study treatment discontinuation?
 - Trying to balance right-sizing TFL volume and value the extra TFL brings
 - The TEAE list is usually quite long, so benefits more for 2 ways of sorting

Adverse Events - FAQs

- Why didn't you include a summary of AEs leading to study (as opposed to study treatment) discontinuation?
 - Disposition tables will include summaries of reasons for discontinuing study treatment and reasons for discontinuing study
 - For specific AEs (eg, MedDRA PTs), only interested in those AEs leading to study treatment discontinue (which is in the recommendations)
 - AEs leading to study discontinuation not as meaningful for safety signal detection

Labs – Treatment Period Shell – Boxplot Change



Endpoint=last postbaseline measure; Box plot type=schematic; The box shows median,interquartile range (IQR, edge of the box), min and max within 1.5 IQR below 25% and above 75% (ends of the whisker). Values outside the 1.5 IQR below 25% and above 75% are shown as outliers. Means plotted as different symbols by treatments. P value is for the treatment comparison from ANCOVA model Change=Baseline+Treatment

Labs - FAQs

- Are measurements collected at an early discontinuation visit considered planned or unplanned?
 - Conceptually, they are planned measurements. However, for labs collected only on occasion, inclusion of such measurements often disrupts the ability to see trends over time.
 - Recommend change to last observation include the measurement
 - Recommend observed and change summaries by visit only keep planned visits per schedule of events

Labs - FAQs

- Why are shifts to high/low summaries based on thresholds as opposed to changes?
 - Conceptually, a case could be made that shift summaries using meaningful changes (instead of thresholds) would be more useful for signal detection for lab data, however they are not available
 - It's unclear whether such an initiative would add sufficient value versus alternative ways of looking at the data

Labs - FAQs

- Instead of boxplots of changes over time, wouldn't analyses of maximum change fit better with the purpose of signal detection?
 - Probably - We may update the recommendations
 - Easier to incorporate a meaningful objective measure of imbalance between treatments
 - Max change is recommended for integrated summaries
 - By visit summaries do not work well for integrated summaries as visit schedules are usually different across studies
 - Having analyses of max changes for integrated, and by visit summaries for individual studies provides a nice overall package

Vitals and ECGs

- Most of what we covered for labs applies to vital signs and quantitative ECGs
 - One exception – shift tables are not recommended when shifts to high/low include a threshold and a change value (which is often done for several vital signs/ECGs)

Patient Profile – Example (showing only 1 of 4 pages)

