AstraZeneca

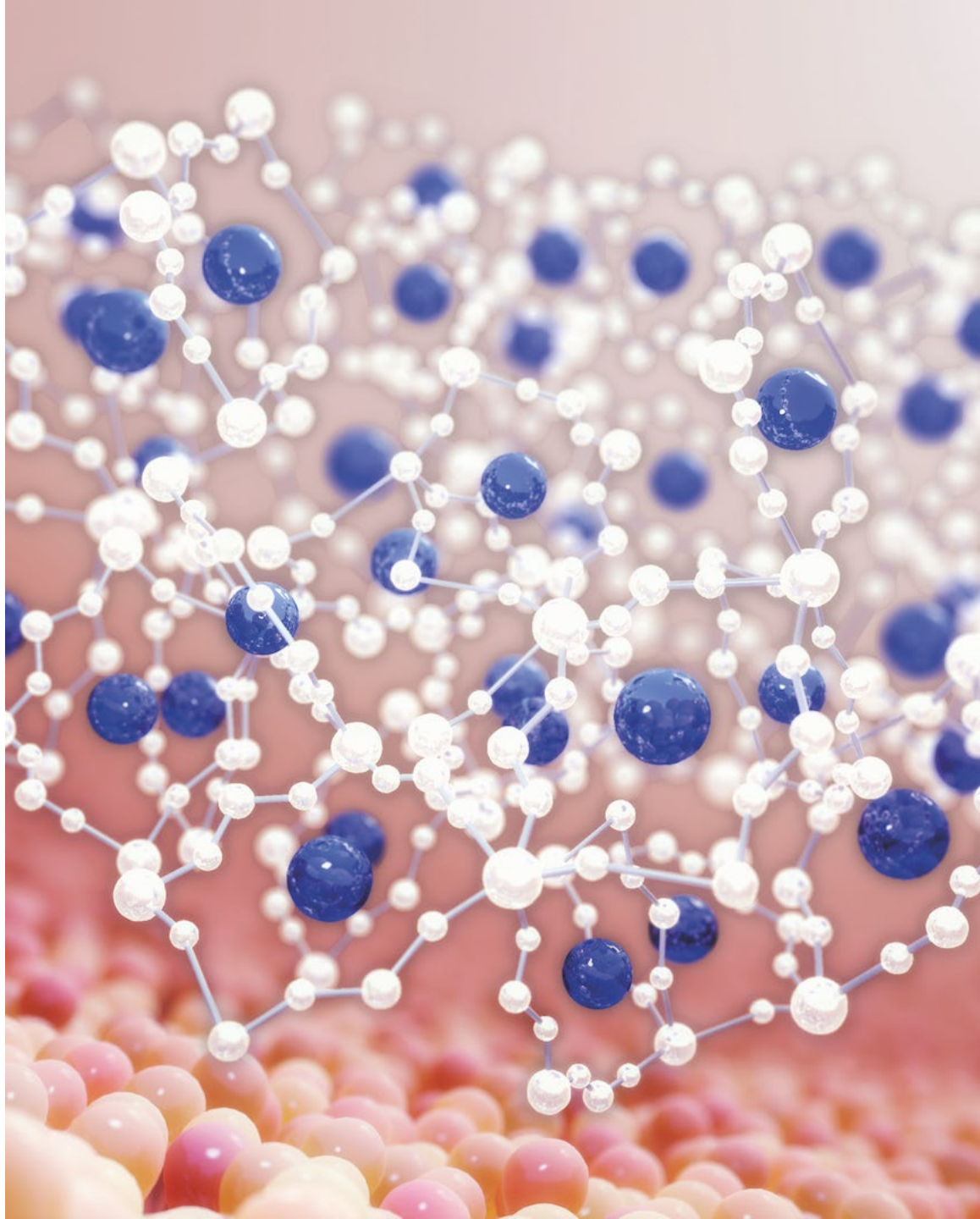
Integrated Analysis of Pre-Marketing Safety Data,

Statistical and Estimand Considerations

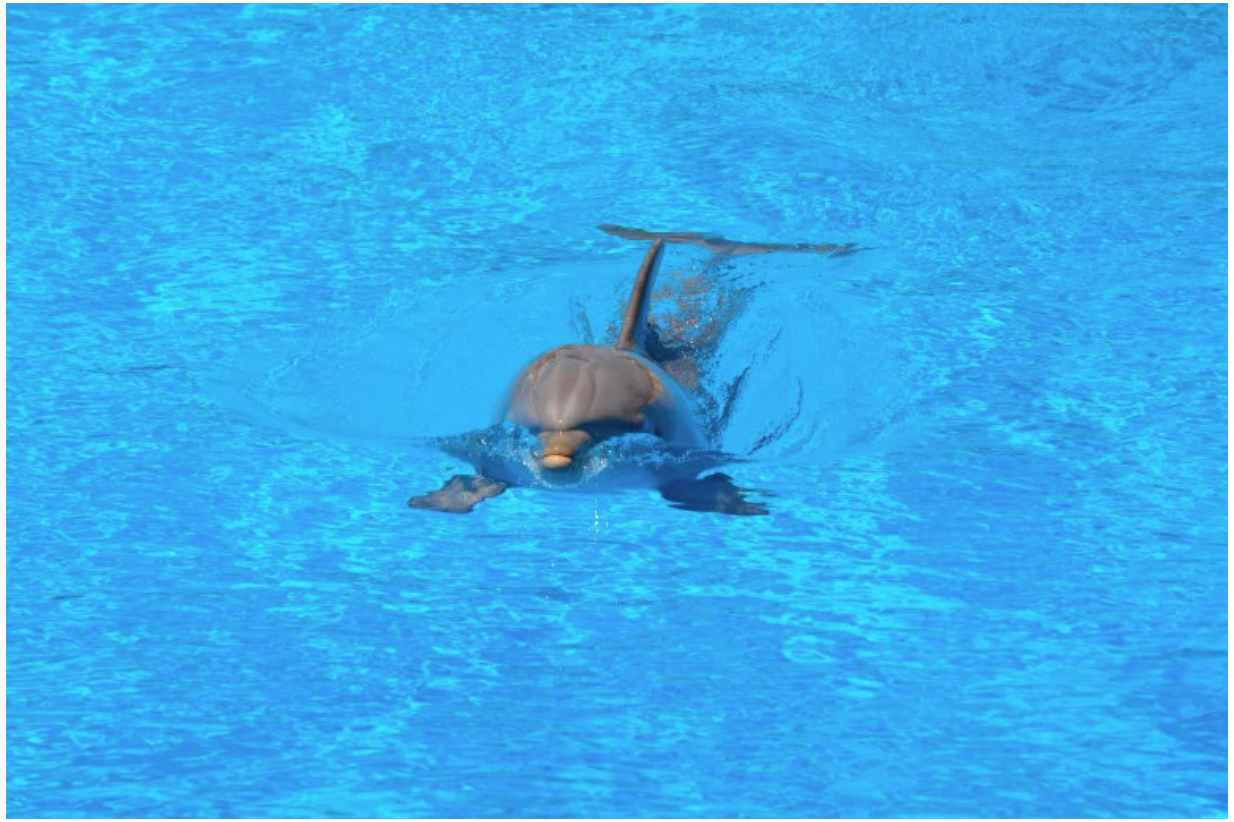
BASS XXXII – 32nd Annual Biopharmaceutical Applied Statistics Symposium

George Kordzakhia (AstraZeneca, Gaithersburg, US)
Katarina Hedman (AstraZeneca, Gothenburg, Sweden)

3 November 2025



Integrated Analysis of Pre-Marketing Safety Data, Statistical and Estimand Considerations



https://pixabay.com/photos/dolphin-swim-water-blue-jump-what-1018685/

PART 1

- What does FDA say
- Statistical Considerations for Aggregated Safety Analysis
- Discussion

PART 2

- Recap of Estimand Fundamentals
- Aggregated Safety Estimands
- Discussion

PHUSE Safety Working Group

This presentation is a draft of an upcoming delivery from Estimands in Safety Analytics Project

phuse.global/working_groups

Acknowledgements

- Hongjian Li, Per Nyström, Andreas Gustavsson, Sharmila Agarwal, Prathamesh Prakash Athavale; Biometrics CVRM, AstraZeneca
- PHUSE Estimands in Safety Analytics Project Team

PART 1

What Does FDA Say?

What does FDA say?

Statistical Considerations for Premarketing Risk Assessment, May 2024

Design and Analysis Plan Principles

- Consider appropriate design and analysis plan elements **prior to Phase 3** to ensure reliable safety evaluation
  - Improved planning improves quality and reliability of safety data
- Appropriate analysis approaches for both AESIs and for descriptive assessment of general safety for signal detection

5

Focus on Estimation and Uncertainty

- Key aspects of safety and benefit-risk assessment
  - Estimation of magnitudes of potential adverse effects
  - Degree of confidence there is not unacceptable harm in context of benefits
- Motivates two goals in safety evaluation
  - Estimation of magnitude of harm (point estimate)
    - Interest in **causal effect**: typically requires randomized comparison unless safety outcome does not occur naturally in population
  - Estimation of uncertainty around magnitude of harm (e.g., confidence interval)
    - Upper CI bound important to consider magnitude of harm ruled out with high confidence

33

G Levin and M Soukup. Statistical Considerations for Premarketing Risk Assessment. FDA workshop May 16, 2024

What does FDA say?

Statistical Considerations for Premarketing Risk Assessment, May 2024

When to Consider Integrated Analyses

- More than one trial collects relevant safety data
- Interest in gaining more precision than provided by individual trials
  - Especially critical for rare events
- Trials are sufficiently similar in design characteristics to assess adverse drug effects
  - Safety outcome definition/ascertainment
  - Appropriateness of exposure and follow-up
  - Appropriateness of comparator and dosing

Appropriate Integrated Analyses

- For a comparison of interest (e.g., drug vs. placebo), typically should include only trials with both treatments
  - May need different trial groupings for different comparisons
- Generally, include only controlled trials/trial periods
  - **CAUTION!** Analyses that include uncontrolled trial periods (e.g., open-label extensions) subject to confounding and bias
- Stratify analyses by trial
  - **CAUTION!** Unstratified analyses of multiple trials may be subject to confounding (see next slide) – Simpson's Paradox
  - **Stratified analyses are always appropriate**

66

G Levin and M Soukup. Statistical Considerations for Premarketing Risk Assessment. FDA workshop May 16, 2024

Statistical Considerations for Aggregated Analysis of Clinical Trial Safety Data

Aggregated Safety Analyses

Safety analyses play a pivotal role in drug development, ensuring the protection of patients while advancing innovative pharmaceuticals to market.

A single study generally does not have sufficient sample size to evaluate all important safety events with reasonable precision and may not cover the full target population for the investigational treatment.

Aggregated (integrated, pooled, meta-) analyses across multiple studies may be helpful in that regard. But without a structured conscious workflow accompanied with appropriate statistical methods for the integrated analysis, this can easily take a route compromising the interpretation.

Hedman, Kordzakhia et al 2024 https://pubmed.ncbi.nlm.nih.gov/39217244

Example. Aggregated Safety Analysis in MASH

Study Design Characteristics	Study 1 - Phase 2B	Study 2 - Phase 3
Treatment Duration	1 year	5 years
Population	MASH Fibrosis Stages: 50% - F2 50% - F3	MASH Fibrosis Stages: 30% - F2 30% - F3 40% - F4
Treatments	Drug A 5 mg Drug A 10 mg Placebo	Drug A 10 mg Placebo
Safety topic of interest	Liver–related adverse events	
Summary measure	Cumulative incidence at time T (T=1, 3, or 5 years)	

- MASH is a chronic progressive liver disease
- The risk of liver related complications increases with progression of liver fibrosis
- Stage F2: Moderate fibrosis
- Stage F3: Advanced fibrosis
- Stage F4: Cirrhosis

Example Aggregated Safety Analysis in MASH

Study Design Characteristics	Study 1 - Phase 2B	Study 2 - Phase 3
Treatment Duration	1 year	5 years
Population	MASH Fibrosis Stages: 50% - F2 50% - F3	MASH Fibrosis Stages: 30% - F2 (Moderate fibrosis) 30% - F3 (Advanced fibrosis) 40% - F4 (Cirrhosis)
Treatments	Drug A 5 mg Drug A 10 mg Placebo	Drug A 10 mg Placebo
Safety topic of interest	Liver–related adverse events	
Summary measure	Cumulative incidence at time T (T=1, 3, or 5 years)	

Can study 1 and 2 be combined to address the following safety questions:

1. Drug A (pool of 5 & 10 mg) vs Placebo in F2+F3+F4 population?
2. Drug A 10mg vs Placebo in F2+F3+F4 population?
3. Drug A 10mg vs Placebo in F2+F3 population?

If yes, to any of the above questions, then at what timepoint(s) can we estimate risk (%)?

Example Aggregated Safety Analysis in MASH

Study Design Characteristics	Study 1 - Phase 2B	Study 2 - Phase 3
Treatment Duration	1 year	5 years
Population	MASH Fibrosis Stages: 50% - F2 50% - F3	MASH Fibrosis Stages: 30% - F2 (Moderate fibrosis) 30% - F3 (Advanced fibrosis) 40% - F4 (Cirrhosis)
Treatments	Drug A 5 mg Drug A 10 mg Placebo	Drug A 10 mg Placebo
Safety topic of interest	Liver–related adverse events	
Summary measure	Cumulative incidence at time T (T=1, 3, or 5 years)	

Conclusion: *Data integration requires careful consideration of targeted population, treatment condition, and summary measure*

Can study 1 and 2 be combined to address the following safety questions:

1. Drug A (pool of 5 & 10 mg) vs Placebo in F2+F3+F4 population?
2. Drug A 10mg vs Placebo in F2+F3+F4 population?
3. Drug A 10mg vs Placebo in F2+F3 population?

If yes, to any of the above questions, then at what timepoint(s) can we estimate risk (%)?

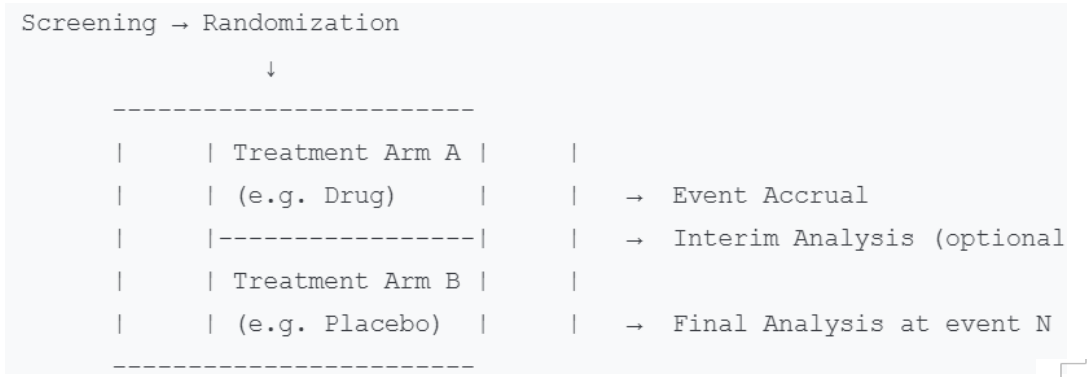
Answers & Considerations:

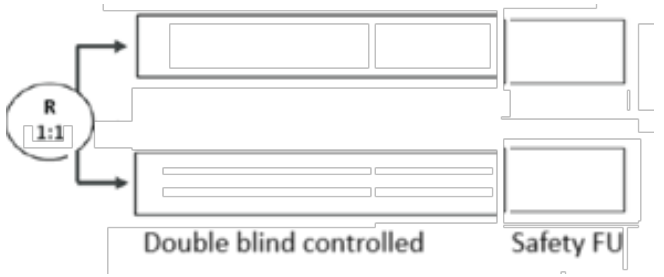
1. No, because 5mg was not evaluated in F4.
2. No. Study 1 does not include F4 participants (F4 is a more severe liver condition). Hence, population attribute in in the aggregated analysis is not well defined. Aggregated treatment differences may not provide a meaningful estimation.
3. ***Yes, but risk (%) estimation should only be based on timepoints $T \leq 1$ year. Crude pooling may potentially lead to biased estimation/Simpson's paradox.***

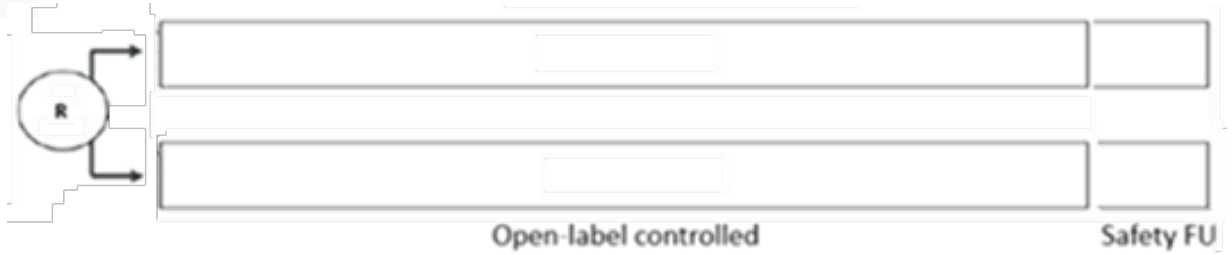
Considerations for Study Aggregation

a. For a given treatment comparison of interest, only the clinical trials that include both relevant treatment arms should be considered for inclusion in the integrated analysis

b. It should be assessed if the candidate studies are double-blind, single-blind, or open-label. As study blinding may have significant impact on the outcome variables, aggregated analysis should not combine blinded and open-label studies







Hedman, Kordzakhia et al 2024 https://pubmed.ncbi.nlm.nih.gov/39217244

Considerations for Study Aggregation

c. In general, *combining different doses is not recommended*, unless a *convincing* supportive rationale is provided, and it is dictated by a clinical question of interest

d. If integration of different active control arms is considered, sound rationale must be provided for the proposed approach and the clinical question to be addressed by this analysis clearly stated

- For example, combining active comparators can be considered if both active comparators serve as a standard of care in clinical practice, belong to the same class, and are known to have similar safety profile

e. Integration of placebo and active arms is generally not recommended

https://pixabay.com/illustrations/pill-capsule-medicine-medical-1884775/

Hedman, Kordzakhia et al 2024 https://pubmed.ncbi.nlm.nih.gov/39217244

Considerations for Study Aggregation

f. If individual studies (will) have different disease or disease severity sub-populations, definition of population attribute requires careful consideration

- For example, if individual studies target subpopulations and the aggregated analysis would better represent the overall population for the indication.
- But, if parameter values being estimated are expected to be *substantially different across populations* in individual trials apriori for the concerned variable, then aggregated analyses may be misleading, or at least not sufficient, and it would be important to evaluate each population individually.

g. Decision to combine studies should also account for how the variable was collected

- For example, a suspected safety signal of heart failure in a phase 2 study may lead to active elicitation or adjudication in phase 3 studies. If so, it may be relevant to perform the analysis separately for each type of ascertainment.

https://pixabay.com/illustrations/mankind-earth-globe-world-people-254788/

Hedman, Kordzakhia et al 2024 https://pubmed.ncbi.nlm.nih.gov/39217244

Study Stratified Aggregated Analysis

https://pixabay.com/illustrations/green-go-sign-road-traffic-1315028/

Use study stratification whenever possible (not simply adding study as covariate to a model applied on pooled subject-level data)

Do not do estimation directly in the pooled subject-level data (meting-pot), without stratifying for study

https://pixabay.com/photos/businessman-mockup-stop-sign-board-3964425/

- May lead to unnecessary bias (even Simpson's paradox)
  - When randomization ratios are different between the studies
  - When studies have different dropout patterns (even if dropout is noninformative within each study) e.g., in KM analyses
- May lead to overestimated variance
  - When studies are enrolling subjects from different geographic regions, different disease sub-populations, or of different degree of severity
  - When aggregating older and more recent studies, since patient management may have changed

Hedman, Kordzakhia et al 2024 https://pubmed.ncbi.nlm.nih.gov/39217244

Study Stratified Aggregated Estimators

Assume k studies are to be utilized to estimate a *probability of occurrence* (P) for a safety event ,

Aggregated estimator $\hat{P}$ will be defined by $\hat{P} = w_1\hat{P}_1 + .. + w_k\hat{P}_k$
where $\hat{P}_i$ are within study estimators, and weights $w_j > 0$ satisfy $\sum w_j = 1$.

Probability of occurrence can be considered if all studies have the same length of the analysis period T.

In a corresponding way, under certain conditions, *aggregated estimates can be constructed also for other common parameters*

- For within treatment group estimation (e.g., incidence rate, mean) and for respective treatment comparisons (e.g., difference or ratios of probabilities, rates, means, odds and hazards)

For some parameters, aggregation of study level estimates is not applicable because the aggregated parameter is not meaningful; for example, percentiles (e.g., minimum, median, maximum).

Hedman, Kordzakhia et al 2024 https://pubmed.ncbi.nlm.nih.gov/39217244

Additional Assumptions for Aggregated Estimation

For aggregated estimation of probability of occurrence by a given timepoint:

- For ***1-KM at timepoint T (%)*** estimator, the aggregated analysis should include only those trials for which reliable KM estimates exist on the interval [0, T].
    - If individual studies have variable treatment duration, T should ideally be set to the median duration of the shortest treatment arm of the shortest study.
- For ***proportion*** estimator, if the studies have fixed but different duration of the analysis period, estimation can only be done on the part of the analysis period that is representative for all included trials.
    - For example, on interval [0, T] where T is the duration of the shortest analysis period of the studies.

Additional Assumptions for Aggregated Estimation

In aggregated analysis some estimators rely on additional assumptions (to what is required for a single study):

For example,

- For estimation of ***Incidence Rate Ratio, Relative Risk (Risk Ratio), Odds Ratio, and Hazard Ratio***
  - Parameter of interest assumed to have the same value across all included studies.

Hedman, Kordzakhia et al 2024 https://pubmed.ncbi.nlm.nih.gov/39217244

Aggregated Estimation

Aggregation of estimators that can allow different values of the parameter of interest across the included studies:

- For estimation of within treatment group parameters
- For estimation of treatment differences
- For example,
    - Proportion (%) and its treatment difference
    - 1-KM (%) and its treatment difference
    - EAIR (per unit of time) and its treatment difference
    - EAER (per unit of time) and its treatment difference

EAIR Exposure adjusted incidence rate, EAER Exposure adjusted event rate; commonly presented by 100 person-years

Hedman, Kordzakhia et al 2024 https://pubmed.ncbi.nlm.nih.gov/39217244

Aggregated 1-KM curve

Do not perform estimation directly in data pooled across studies at a subject level

Easy to produce a 1-KM curve for the study stratified analysis

1. *Within each study*, calculate 1-KM at all timepoints where an event happens
2. Calculate *study stratified aggregated 1-KM* at every timepoint an event happens in any of the studies

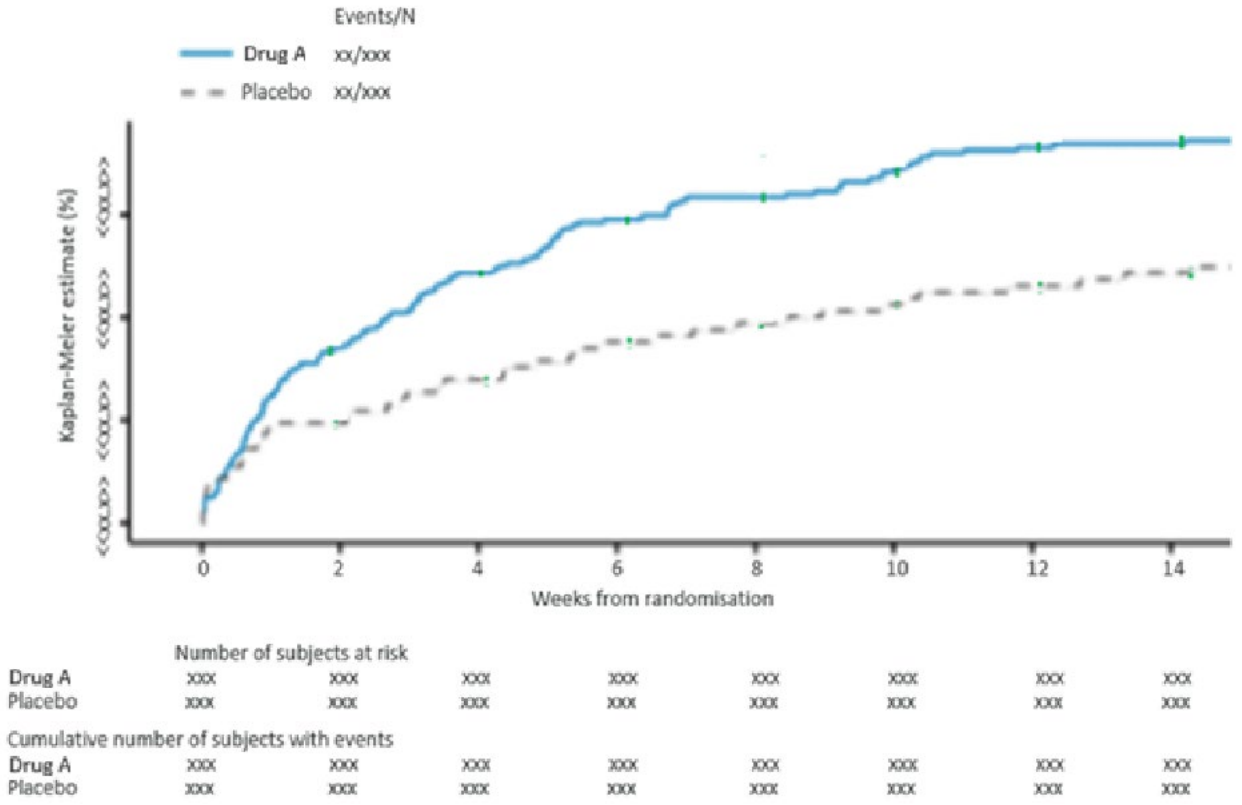


Figure is based on US Food and Drug Administration, standard safety tables and figures: integrated guide, August 2022; https://www.regulations.gov/document/FDA-2022-N-1961-0046

1-KM = 1 minus Kaplan-Meier survival estimate

Hedman, Kordzakhia et al 2024 https://pubmed.ncbi.nlm.nih.gov/39217244

What Study Weights?

https://pixabay.com/illustrations/green-go-sign-road-traffic-1315028/

Several possible weighting options can be considered, for example

- Weighting by the *study sample size* is commonly generally applicable
- Weighting based on estimand related considerations
  - For example, if individual studies enroll subjects from *different subpopulations*, study weights can be chosen to reflect representation of these subpopulations in the targeted disease population for the experimental drug

For survival estimates or crude risk (incidence proportion)

- *Avoid using* the common meta-analysis method of *weighting by inverse of variance*
  - Implies a dependency between the study weights and the study outcome of risk estimate
  - Different study weights when performing integrated analysis of risk at different post-baseline timepoints
  - Different study weights for different types of safety events

https://pixabay.com/photos/businessman-mockup-stop-sign-board-3964425/

Hedman, Kordzakhia et al 2024 https://pubmed.ncbi.nlm.nih.gov/39217244

Study Size Proportional Weights Example

Aggregated analysis sets	Number of participants				Study Weight[1]	
	Drug A 5 mg	Drug A 10 mg	Placebo	Total	Formula	Proportion
ISS set 5 mg						
Study 1	401		398	799	799/1202	0,6647
Study 2	198		205	403	403/1202	0,3353
Sum across studies	599		603	1202	1	1
ISS set 10 mg						
Study 1		395	398	793	793/1409	0,5628
Study 2		203	205	408	408/1409	0,2896
Study 3		106	102	208	208/1409	0,1476
Sum across studies		704	705	1409	1	1

In this example, focusing the aggregated analyses on Drug A dose X vs Placebo

- For each aggregated analysis set, the *study weights are preferably kept constant* for all aggregated safety analyses.
- Implying that for a *subgroup analysis* (eg females) of an aggregated analysis set (eg ISS set 5 mg), the unique study weight for that complete analysis set will be used (eg study 1 will use weight 0.6647).
- This provides consistency in population attribute across strata and enables meaningful interpretation of estimates across subgroup strata as well as treatment comparisons within a subgroup stratum.

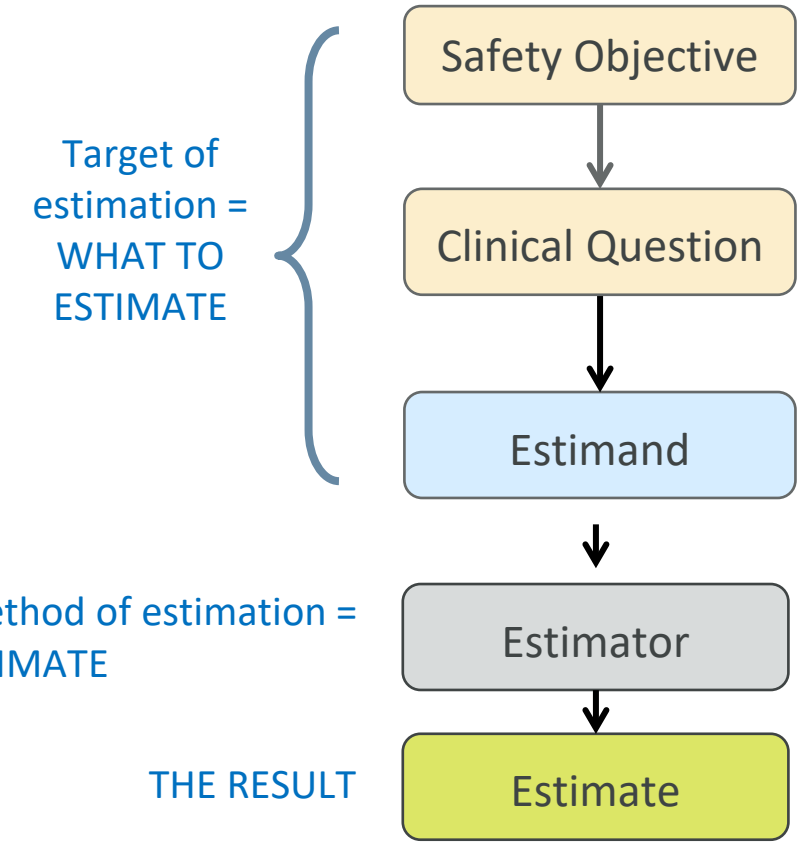
[1]PHUSE 2017 AE


Q & A

PART 2

Recap of Estimand Fundamentals

Estimand, Estimator, Estimate



E.g., the general safety objective: To evaluate the safety and tolerability of the investigational treatment.

E.g., what is the magnitude of the increase in risk (%) of experiencing an adverse event 1 year after the treatment is initiated, for drug A as compared to placebo in persons with diabetes?[1]

The "recipe" for obtaining an answer to the clinical question.

E.g., within treatment group incidence proportion and its treatment difference.

E.g., 10%, as in the incidence proportion for an adverse event was 10% higher in drug A group compared to placebo.

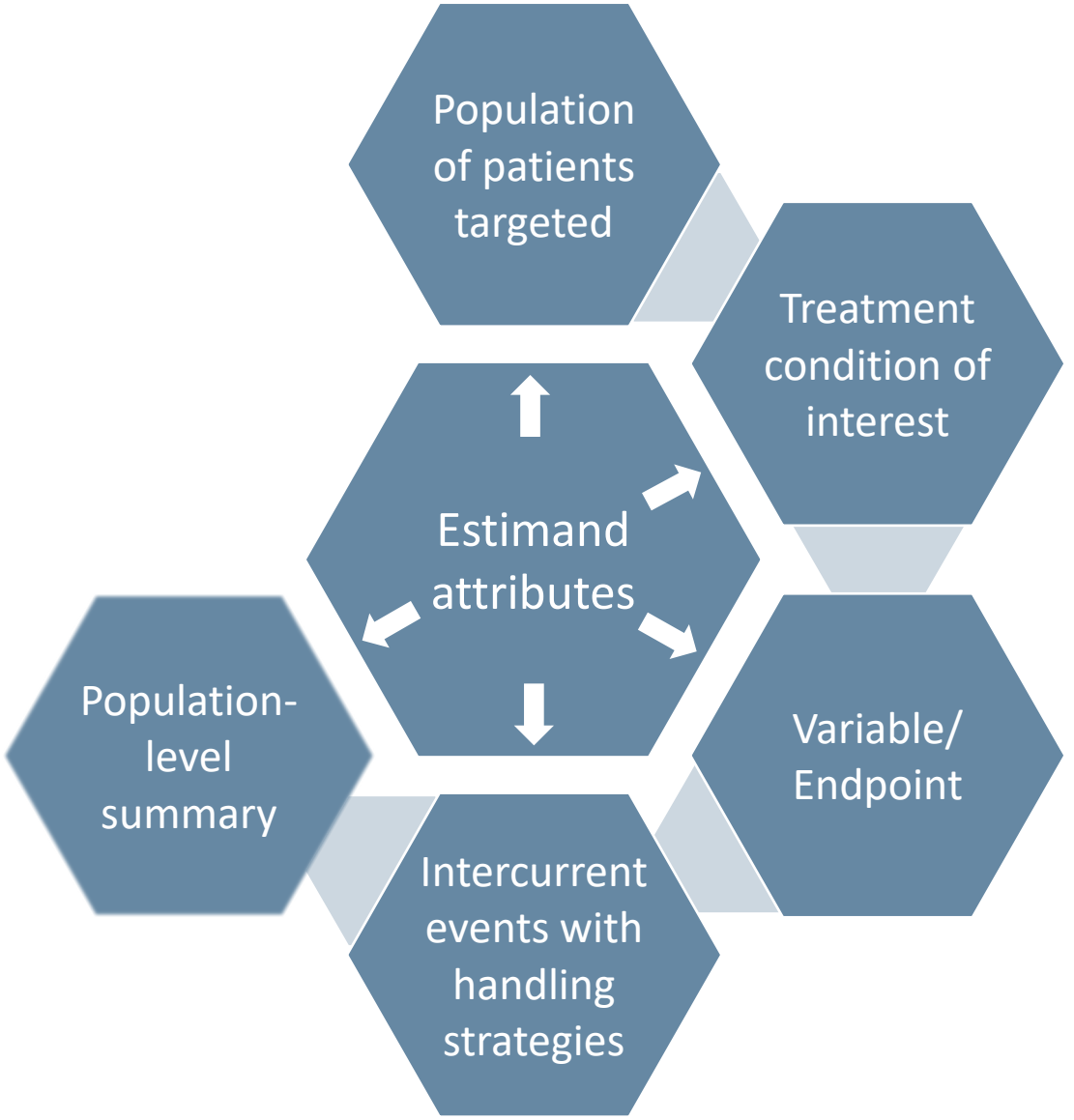
ICH E9(R1) guidance and training, PHUSE safety working group Estimands

What is an Estimand?

The Estimand is the target of estimation to address a specific clinical question. It is commonly described by five attributes**.**

Intercurrent events are events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest.

- E.g., premature treatment discontinuation, change in background medication and death.



PHUSE safety working group Estimands , ICH E9(R1), guidance and training, https://cddf.org/wp-content/uploads/2020/12/CDDF_Estimands_Speaker-Lecture_2020-12-01_Final-Version.pdf

Common Estimand Attributes

POPULATION	Persons for whom researchers want to evaluate the safety profile
TREATMENT CONDITION	Defined through a particular dosing regimen of an investigational intervention.
VARIABLE (ENDPOINT)	Obtained for *each participant*, e.g., an indicator if the defined event has happened or not or a laboratory measurement.
INTERCURRENT EVENTS with handling strategies	Key intercurrent events in safety analysis are commonly premature treatment discontinuation, change in background medication and death.
POPULATION-LEVEL SUMMARY (SUMMARY MEASURE)	A *parameter* that summarizes the distribution within treatment group or for a comparison between treatment groups. - E.g., probability of an event occurring for the participants in the treatment group, or treatment difference in the probabilities.

An appropriate **ESTIMATOR** is then used to produce the summary measure **ESTIMATE**

ICH E9(R1), guidance and training, PHUSE safety working group Estimands

Common Intercurrent Event Handling Strategies

TREATMENT POLICY The ICE is effectively ignored in the analysis (safety event must be observable after the ICE)

HYPOTHETICAL Envisages a scenario in which the ICE would not occur

WHILE-ON-TREATMENT or WHILE-ALIVE Used when response to treatment before the occurrence of the ICE (treatment discontinuation or death) is of interest, irrespectively when the ICE happens

COMPOSITE Used when the ICE is considered to be informative about the participant's endpoint and is therefore incorporated into the definition of the endpoint

ICE Intercurrent Event

ICH E9(R1), guidance and training, PHUSE safety working group Estimands

Estimands for Aggregated Analysis

Aggregated Analysis Require Clear Clinical Questions

- All analyses correspond to some estimands, but if these estimands are not clearly stated (are hidden) we may not know if they are meaningful or relevant to the intended clinical questions.

https://pixabay.com/photos/cat-black-cat-cat-eyes-animal-4043419/

- *Take control of the aggregated evaluation* by defining clear clinical questions and match each key question to a clearly defined meaningful estimand.

PHUSE safety working group Estimands

Typical Clinical Questions

1. ***Signal detection:***
   What is the magnitude of the increase in *risk (%)* of experiencing a safety event *up to 1 year after initiation of treatment* for Treatment A compared to Placebo in adults with type 2 diabetes?

2. ***Signal detection alternative:***
   What is the magnitude of the increase in *incidence rate (per time unit)* of experiencing a safety event *during treatment* with Treatment A compared to Placebo in adults with type 2 diabetes?

3. ***Assessment of a Late Latency AESI*** *(alternative or supplementary to question 1):*
   What is the magnitude of the increase in *incidence rate (per time unit)* of experiencing bone fracture *after initiating treatment* with Treatment A compared to Placebo in adults with type 2 diabetes?

4. ***Assessment of an AESI in Study With Higher Mortality*** *(competing risks, supplementary to question 1):*
   What is the magnitude of the increase in *risk (%)* of experiencing bone fracture *before premature treatment discontinuation or death for up to 1 year after initiation of treatment*, for Treatment A compared to Placebo in adults with type 2 diabetes?

PHUSE safety working group Estimands

Estimands for Aggregated Safety Analyses

Estimand attributes	Considerations for estimand for Aggregated Analysis (across multiple studies) versus corresponding estimand for an Individual study
Population	The aggregated analysis tends to represent a broader population than the population for each included individual study.
Treatment condition	Treatment condition in aggregated analysis may differ from those in the individual studies. Without proper justification, pooling different dose regimens or pooling active control and placebo may confound the interpretation.
Endpoint/ Variable	The variable selected for the integrated analysis should be assessed in each of the individual studies. For example, lab values should be assessed at similar timepoints
Intercurrent events & strategies	Key intercurrent events are expected to be the same in aggregated analysis and in individual studies but can be population and dose specific. For each type of intercurrent event, the strategy selected for the aggregated analysis should be consistent with the clinical question.
Summary measure	The population-level summary should be the same, and evaluable in each study. However, the aggregated analysis may require additional assumptions.

Hedman, Kordzakhia et al 2024 https://pubmed.ncbi.nlm.nih.gov/39217244

Examples of Aggregated Analysis Estimands

Example 1. Aggregated Safety Analysis in MASH

Study Design Characteristics	Study 1 (Phase 2B)	Study 2 (Phase 3 and 4)
Treatment Duration	1 year	5 years
Population	MASH Fibrosis Stages 2 & 3	MASH Fibrosis Stages: 60% 2 & 3, and 40% 4 (cirrhosis)
Treatments	Drug A 5 mg, Drug A 10 mg, or Placebo	Drug A 10 mg, or Placebo
Safety topic of interest	Liver–related adverse events	

Study 1 and 2 can for example be combined to address clinical questions:

1a. What is the magnitude of the increase in *risk (%)* of experiencing a liver-injury adverse event *up to 1 year after initiation of treatment* for *Drug A 10 mg* compared to Placebo in adults with *MASH fibrosis Stages 2 & 3*?

1b. What is the magnitude of the increase in *incidence rate* of experiencing a liver-injury adverse event *during treatment* with *Drug A 10 mg* compared to Placebo in adults with *MASH fibrosis Stages 2 & 3*?

Example 1a. Aggregated Safety Analysis in MASH

1a. Clinical Question for Aggregated analysis:
What is the magnitude of the increase in *risk (%)* of experiencing a liver-injury adverse event *up to 1 year after initiation of treatment* with *Drug A 10 mg* compared to Placebo in adults with *MASH fibrosis Stages 2 & 3*?

Population: Adults with MASH fibrosis Stages 2 & 3.

Treatment condition: Drug A 10 mg or Placebo.

Variable: A combination of a 'time to' a liver-injury AE and time to censor.

Main summary measures: *Probability of occurrence before 1 year* and its treatment difference, estimated by '1 minus Kaplan-Meier survival' *(1-KM)* estimator.

Key intercurrent events (ICEs) with handling strategies:
- Ignoring treatment discontinuation and changes in background medications *(Treatment policy strategy)*; requiring data collection continues after the ICEs.
- Not imputing any data after death and using 1-KM estimator *(Hypothetical strategy)*.

Example 1b. Aggregated Safety Analysis in MASH

1b. Clinical Question for Aggregated analysis:
What is the magnitude of the increase in *incidence rate* of experiencing a liver-injury adverse event *during treatment* with *Drug A 10 mg* compared to Placebo in adults with *MASH fibrosis Stages 2 & 3*?

Aggregated Analysis Estimand:

Population: Adults with MASH fibrosis Stages 2 & 3.

Treatment condition: Drug A 10 mg or Placebo.

Variable: 'Time to' a liver-injury AE.

Main summary measures: *Incidence rate per unit of time* and its treatment difference implemented by *EAIR* per 100 person-years.

Key intercurrent events (ICEs) with handling strategies for the main estimators:
- Censoring the data at 2 weeks after treatment discontinuation and using EAIR estimator (*While-on-treatment strategy*).
- Not imputing any data after death and using EAIR estimator (*While-alive strategy*).
- Ignoring changes in concomitant medications (*Treatment policy strategy*).

Example 2a. Aggregated Safety Analysis in population A+B

Study Design Characteristics	Study 1	Study 2	Study 3
Treatment Duration	1 year	1 year	2 years
Population	A	B	A+B
Treatments	drug X 5mg, drug X 10mg, placebo	drug X 5mg, drug X 10mg, placebo	drug X 10mg, placebo
Randomization Ratio	2:2:1	1:1:1	1:1

Clinical Question for Aggregated analysis:
What is the magnitude of the increase in *risk (%)* of experiencing a safety event *up to 1 year after initiation of treatment for drug X 5 mg* compared to placebo in mixed population of A and B?

a) Pre-specify appropriate weights w_1 *and* w_2 *to aggregate populations A and B* (e.g., based on representation in the target population for the experimental treatment).

b) Aggregate Study 1 and Study 2 estimates for 5mg, placebo, and treatment difference, respectively, using w_1 *and* w_2 to obtain estimates in the mixed population.

Note 1. Utilizing fixed weights for aggregation addresses potential issues with different randomization ratios in Study 1 and Study 2.

Note 2. The same approach can be used to compare 5mg with 10mg (if it is of interest). Using the same weights for aggregation of A and B in the three treatment groups ensures comparability of the population attribute.

Weights for populations must satisfy:

$$w_1 > 0, w_2 > 0$$

$$w_1 + w_2 = 1$$

Example 2b. Aggregated Safety Analysis in population A+B

Study Design Characteristics	Study 1	Study 2	Study 3
Treatment Duration	1 year	1 year	2 years
Population	A	B	A+B
Treatments	drug X 5mg, drug X 10mg, placebo	drug X 5mg, drug X 10mg, placebo	drug X 10mg, placebo
Randomization Ratio	2:2:1	1:1:1	1:1

Clinical Question for Aggregated analysis:
What is the magnitude of the increase in *risk (%)* of experiencing a safety event *up to 1 year after initiation of treatment for drug X 10 mg* compared to placebo in mixed population of A and B?

a) Pre-specify appropriate weights w_1 *and* w_2 *for aggregation of populations A and B* (e.g., based on representation in the target population for the experimental treatment).

b) Aggregate Study 1 and Study 2 estimates for 10mg, placebo, and treatment difference, respectively, using weights w_1 *and* w_2.

c) In Study 3, perform estimation for 10mg, placebo, and treatment difference in sub-populations A and B. Aggregate estimates across sub-populations A and B using the same weights as in step b.

d) Pre-specify weights v_1 *and* v_2 to integrate Study 3 with Studies 1 and 2. Aggregate the aggregated estimates from steps b and c, *using* v_1 *and* v_2. Weights can for example be based on the total sample sizes in Study 1+2 and Study 3, for the included treatments (10 mg and placebo).

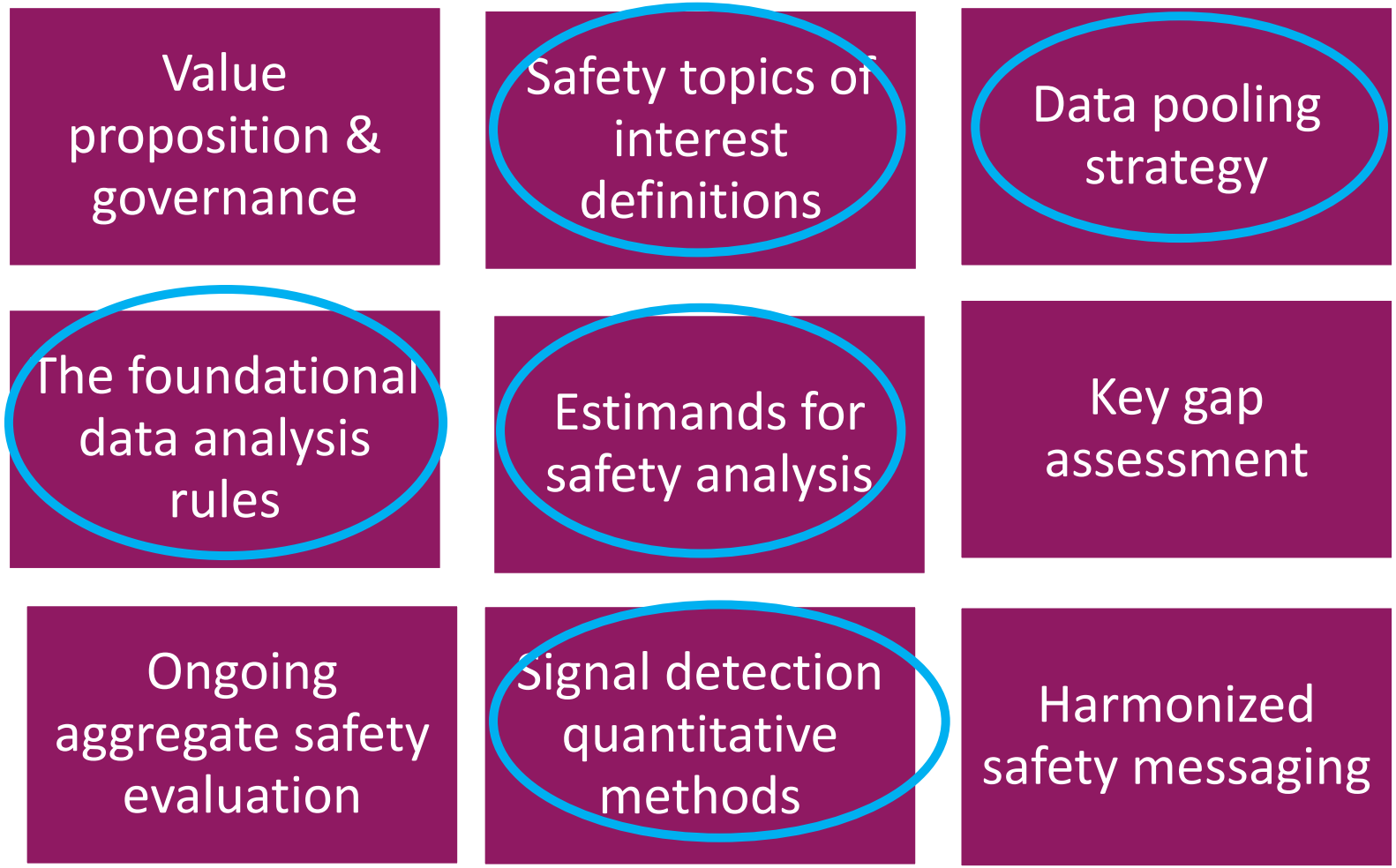
Weights for aggregation must satisfy:

$$v_1 > 0,\ v_2 > 0$$

$$v_1 + v_2 = 1$$

Aggregated Analysis Require More Planning
(than corresponding analysis in a single study)

Aggregate Safety Assessment Planning (ASAP) modular framework for a drug development program



- The Biometrics modules, empowers interpretation of aggregated safety analyses across multiple studies
- Frontloading workload for project team

Ch 2 in Wang, Munsaka et al, Quantitative drug safety and benefit risk evaluation – practical and cross-disciplinary approaches 2021
Details in Hendrickson, Wang, Ball et al, Aggregate Safety Assessment Planning for the Drug Development Life-Cycle, TIRS (2021) 55:717–732, https://pubmed.ncbi.nlm.nih.gov/33755928/

Aggregated Evaluation Requires More Planning

Early in the drug project, preferably in an ASAP modular framework for a drug development program[1]

https://pixabay.com/vectors/list-icon-symbol-paper-sign-flat-2389219/

1. Define the key clinical questions
   - Define what pair of treatments to compare in what population
2. Define criteria for what type of studies to aggregate per each key clinical question
3. Update the above when required

For the individual studies considered for the aggregated evaluation, what can be made similar across the studies to empower the aggregated analysis, without threatening the aim of the individual study and without adding unreasonable costs?[2]

ASAP Aggregate Safety Assessment Planning

1) Hendrickson, Wang and Ball et al 2021 https://pubmed.ncbi.nlm.nih.gov/33755928/ ; 2) Hedman, Kordzakhia et al 2024 https://pubmed.ncbi.nlm.nih.gov/39217244

Aggregated Evaluation Requires More Planning

SAP writing for the defined Aggregated Evaluation

Preferably complete the SAP of the aggregated evaluation prior 'first subject in' in the studies to be included (e.g., SAP for an ISS for a submission to FDA)

https://pixabay.com/vectors/list-icon-symbol-paper-sign-flat-2389219/

4. Select the clinical questions

5. Select the studies that apply to the inclusion criteria for the aggregated evaluation, the study phases (e.g., controlled treatment phase and safety follow-up (if it follows seamlessly after the controlled treatment phase)) and treatment groups to include.

6. Decide timing of commencing of the aggregated evaluation
   - For example, clinical data lock in study X.

7. Define an Estimand for each key clinical question
   - Population, Treatment condition, Variable, Summary measure, Intercurrent events with handling strategies

8. Define how to implement the estimands
   - Analysis sets (both for describing the included studies and for the safety analyses), Study weights, Estimators etc.

SAP Statistical Analysis Plan


Q & A

Extra Topics

Multiplicity in Safety Analysis

- The safety analysis may be evaluating thousands of preferred terms, and hence, the chance of an erroneous finding could be substantial.
- Multiplicity issues primarily come from the very large number of variables that are evaluated, another source of multiplicity comes from repeated assessment of the same safety parameter over time.
- Mitigation
  - SAP text: All safety analyses for general safety objective are descriptive, no p-values will be presented, and the confidence intervals are provided as illustration of precision of the estimates, and their coverage should not be interpreted as a proof of an existence, or an absence, of a safety signal.
  - Patient safety adds medical, pharmacological and chemical compound knowledge and identify clinical concepts, to arrive to a conclusion of an ADR.

Table 1. System organ class cardiac disorders by preferred term – on-treatment analysis period (full analysis set).

System organ class preferred term	Drug A Number of subjects = 96		Control Number of subjects = 104		Difference in Kaplan–Meier estimate % (95% confidence interval)[a] at 14 months Drug A versus control	Hazard ratio[b] (95% confidence interval) Drug A versus control
	Number of subjects with event (%)	Kaplan-Meier estimate % at 14 months	Number of subjects with event (%)	Kaplan-Meier estimate % at 14 months		
Cardiac disorders	18 (18.8)	19.5	28 (26.9)	28.5	−9.0 (−21.0, 3.1)	0.65 (0.36, 1.18)
Acute myocardial infarction	0	0	1 (1.0)	1.1	−1.1 (NC)	0 (NC)
Angina pectoris	1 (1.0)	1.0	0	0	1.0 (NC)	Infinity (NC)
Angina unstable	1 (1.0)	1.0	1 (1.0)	1.1	−0.1 (−3.0, 2.9)	1.04 (0.06, 16.58)
Arrhythmia	0	0	1 (1.0)	1.0	−1.0 (NC)	0 (NC)
Atrial fibrillation	4 (4.2)	4.5	3 (2.9)	3.2	1.3 (−4.2, 6.9)	1.37 (0.31, 6.13)
Atrioventricular block	1 (1.0)	1.1	0	0	1.1 (NC)	Infinity (NC)
Bradycardia	1 (1.0)	1.1	0	0	1.1 (NC)	Infinity (NC)
Cardiac arrest	1 (1.0)	1.2	0	0	1.2 (NC)	Infinity (NC)
Cardiac failure	8 (8.4)	8.8	17 (16.3)	17.7	−8.9 (−17.8, 1.0)	0.54 (0.24, 1.21)

Hedman K, Lisovskaja V, Nyström P. A safety estimand for late phase clinical trials where the analysis period varies over the subjects. Clin Trials. 2024;21(4):483–490.

CI for Study Stratified Aggregated Estimator

Within treatment group variability and confidence intervals for survival function

Suppose that we form a "combined" population P given by a weighted mixture of populations P_i with weights $w_i > 0,\ i = 1,..,k.$, survival function $S(t)$ in the "combined" population P can be represented as a linear combination of survival functions $S_i(t)$:

$$S(t) = \sum_{i=1}^{k} w_i\, S_i(t)$$

For a weighted linear combination $\hat{S}(t) = \sum_{i=1}^{k} w_i\, \hat{S}_i(t)$, noting that $\hat{S}_i(t),\ i = 1,..,k$ are independent, the variance can be written as:

$$\hat{\sigma}^2(t) = \sum_{i=1}^{k} w_i^2\, \hat{\sigma}_i^2(t).$$

Kaplan-Meier estimator at time t $(t > 0)$ is approximately normally distributed for large sample sizes. The same would hold for weighted linear combinations. The 95% confidence interval is given by

$$(\,\hat{S}(t) - z_{0.975} \cdot \hat{\sigma}(t)\,,\ \hat{S}(t) + z_{0.975} \cdot \hat{\sigma}(t)\,)$$

For the cumulative incidence function estimator $\hat{I}(t) = 1 - \hat{S}(t)$, variances are identical to those of $\hat{S}(t)$. Respective 95% confidence interval for $I(t)$ will be:

$$(1 - \hat{S}(t) - z_{0.975} \cdot \hat{\sigma}(t)\,,\ 1 - \hat{S}(t) + z_{0.975} \cdot \hat{\sigma}(t)\,)$$

Hedman, Kordzakhia et al 2024 https://pubmed.ncbi.nlm.nih.gov/39217244; E. L. Kaplan and Paul Meier, Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association, Vol. 53, No. 282 (Jun 1958), pp. 457-481; Thomas R. Fleming and D.Y. Lin, Survival Analysis in Clinical Trials: Past Developments and Future Directions. Biometrics 56, 971-983, December 2000.

CI for Study Stratified Aggregated Estimator

Treatment comparisons and confidence intervals

The difference of survival curves between the investigational arm (T) and the comparator (C), can be estimated by

$\hat{S}_T(t) - \hat{S}_C(t)$

The variance of the difference is the sum of variances $\hat{\sigma}_T{}^2(t) + \hat{\sigma}_C{}^2(t)$, and thus, the confidence interval is given by:

$$\left(\left(\hat{S}_T(t) - \hat{S}_C(t)\right) - z_{0.975} \cdot \sqrt{\hat{\sigma}_T{}^2(t) + \hat{\sigma}_C{}^2(t)}\,,\ (\hat{S}_T(t) - \hat{S}_C(t)) + z_{0.975} \cdot \sqrt{\hat{\sigma}_T{}^2(t) + \hat{\sigma}_C{}^2(t)} \right)$$

The difference of cumulative incidences can be written as

$$I_T(t) - I_C(t) = \left(1 - S_T(t)\right) - \left(1 - S_C(t)\right) = -\left(S_T(t) - S_C(t)\right)$$

Hence, the estimate for the difference and corresponding confidence interval can be obtained by multiplying respective estimates for $S_T(t) - S_C(t)$ by -1. The respective confidence interval will be:

$$\left(-\left(\hat{S}_T(t) - \hat{S}_C(t)\right) - z_{0.975} \cdot \sqrt{\hat{\sigma}_T{}^2(t) + \hat{\sigma}_C{}^2(t)}\,, -(\hat{S}_T(t) - \hat{S}_C(t)) + z_{0.975} \cdot \sqrt{\hat{\sigma}_T{}^2(t) + \hat{\sigma}_C{}^2(t)} \right)$$

Hedman, Kordzakhia et al 2024 https://pubmed.ncbi.nlm.nih.gov/39217244; E. L. Kaplan and Paul Meier, Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association, Vol. 53, No. 282 (Jun 1958), pp. 457-481; Thomas R. Fleming and D.Y. Lin, Survival Analysis in Clinical Trials: Past Developments and Future Directions. Biometrics 56, 971-983, December 2000.