U.S. FOOD & DRUG ADMINISTRATION

Estimand and Missing Data

Annual Biopharmaceutical Applied Statistics Symposium (BASS)

Abel Eshete
Lead Mathematical Statistician
Division of Biometrics IV, Office of Biostatistics
OTS/CDER/FDA

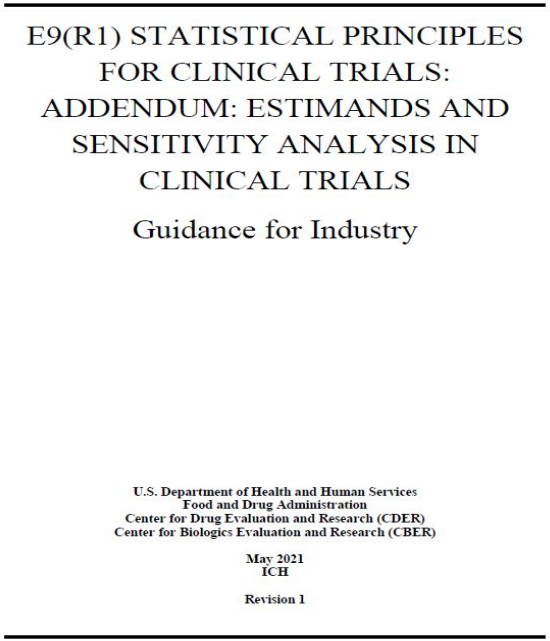
Disclaimer

- The views expressed in this presentation are those of the author and should not be construed as representing the official position of the U.S. Food and Drug Administration.

Part 1: Estimands

- Defining estimands
- Why do we need estimands?
- Attributes of an estimand
- Intercurrent events
- Intercurrent events strategies

What is an "*Estimand*"?



E9(R1) STATISTICAL PRINCIPLES FOR CLINICAL TRIALS: ADDENDUM: ESTIMANDS AND SENSITIVITY ANALYSIS IN CLINICAL TRIALS

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2021
ICH

Revision 1

https://www.fda.gov/media/148473/download

A precise description of the treatment effect reflecting the clinical question posed by the trial objective. It summarizes at a population-level what the outcomes would be in the same patients under different treatment conditions being compared."

Why Do We Need Estimands?

- Before ICH E9(R1), clinical trials often described the treatment effect only implicitly — buried inside the choice of statistical method.

- This created ambiguity:

  – What population does the effect apply to?
  – What happens to a patient who drops out or switches therapy?
  – Are two trials even answering the same clinical question?

- ICH E9(R1) (adopted November 2019; FDA guidance finalized 2021) introduced the estimand framework to ensure:

  – Alignment between clinical question, study design, data collection, and analysis
  – Greater precision and transparency in describing the treatment effect

- Think of the estimand as the 'target' — you must define what you are aiming at before you can measure how close you got

Attributes of an Estimand

- An estimand has FIVE attributes:
  - Treatment: what is being compared (test vs. control)
  - Population: which patients does the question target?
  - Variable (Endpoint): what outcome is measured for each patient?
  - Population-level summary: how is the endpoint aggregated & compared?
  - Intercurrent events (ICEs): how are post-randomization events handled?
- Defining the estimand first drives everything else: design, data collection, statistical analysis, and interpretation.

Intercurrent Events (ICEs)

- **Definition**: Events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest. — ICH E9(R1)

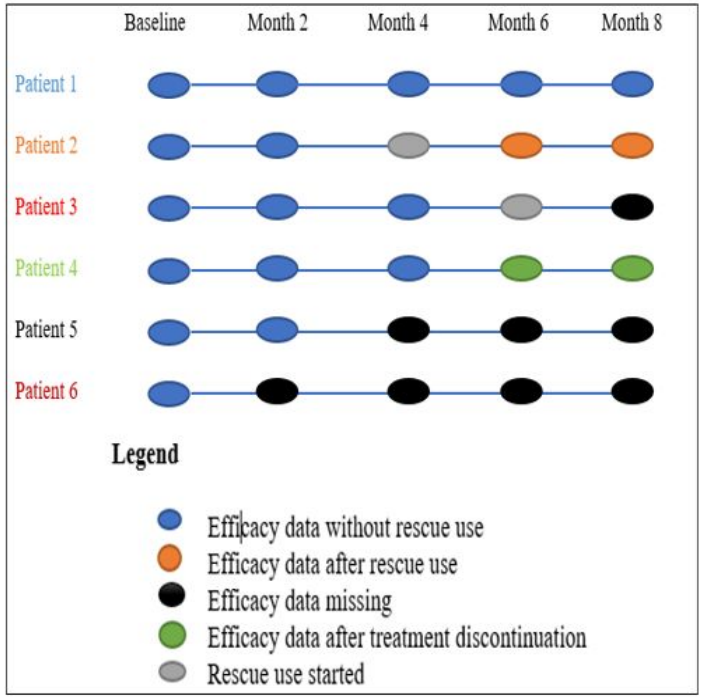
- Common ICEs in clinical trials:
  - Discontinuation of assigned treatment (AE, lack of efficacy, patient choice)
  - Use of prohibited or rescue medication
  - Death before the primary endpoint timepoint
  - Treatment switch

- Key distinction to keep in mind:
  - ICE occurred → the desired data do not exist or are not meaningful due to the event
  - Missing data → the meaningful data exist but were not collected

- The strategy chosen to handle each ICE determines what 'missing data' means for the analysis.

Patient Experiences

- Consider an 8-month study where patients are evaluated every other month.



- **Patient 1** represents subjects who completed the study without treatment discontinuation or a need for rescue medication and have efficacy outcome at all time points.
- **Patient 2** represents subjects who completed the study and provided efficacy outcome at all months. However, because a rescue medication is used at Month 4, the efficacy outcomes at the subsequent months are affected by the rescue medication.
- **Patient 3** represents subjects who received a rescue medication and discontinued the study thereafter. Consequently, an efficacy outcome at Month 8 is missing.
- **Patient 4** represents subjects who discontinued the study treatment due to toxicity/lack of efficacy but remained in the study and provided efficacy data at subsequent months.
- **Patient 5** represents subjects who discontinued the study treatment due to lack of efficacy/toxicity and left the study without providing efficacy data at subsequent months.
- **Patient 6** represents subjects who are lost to follow up and therefore, efficacy data is missing for all visits after Month 2.

Intercurrent Strategies

- ICH E9(R1) outlines five main strategies. The choice should be driven by the clinical question — not statistical convenience.
  - Treatment Policy
  - Composite Variable
  - Hypothetical
  - While-on-Treatment
  - Principal Stratum
- One or a combination of these strategies may be used to address a given estimand.

Intercurrent Strategies

- **Treatment Policy Strategy**
  - **Objective**: Estimate the treatment effect regardless of whether patients adhered to, discontinued, or switched treatment — all post-ICE data are included as observed.
  - **How it works**: All subjects are analyzed using their full observed data, irrespective of ICE occurrence. Conceptually aligned with the ITT principle.
  - **Advantage**: Reflects realistic treatment use in a population where non-adherence, rescue, and switching are possible.
  - **Limitation**: If many patients leave the study and are unavailable for follow-up, this strategy becomes difficult to execute and may introduce substantial missing data.

Intercurrent Strategies

- **Hypothetical Strategy**
  - **Objective**: Estimate the treatment effect under a scenario in which the ICE never occurred — For example, what the outcome would have been had patients remained on treatment without interruption.
  - **How it works**: Post-ICE data are censored, and the missing outcome is estimated via statistical modeling (e.g., multiple imputation, MMRM) under pre-specified assumptions.
  - **Advantage**: Provides an estimate of the pharmacological effect under idealized conditions.
  - **Limitation**: Relies on untestable assumptions about the counterfactual post-ICE trajectory. Sensitivity analyses are essential, and results become increasingly assumption-driven as ICE rates rise.

Intercurrent Strategies

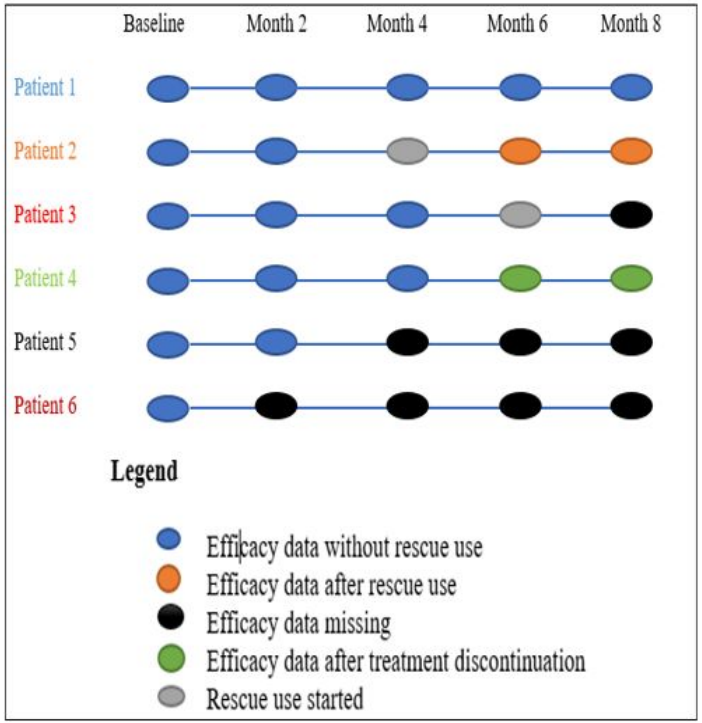
- **Composite Variable Strategy**
  - **Objective**: Evaluate the treatment effect by incorporating the ICE directly into the definition of the efficacy outcome — the occurrence of the ICE is part of the result.
  - **How it works**: Subjects who experience the ICE are assigned a pre-specified unfavorable outcome value — for example, for binary endpoints, they are counted as non-responders/failures.
  - **Advantage**: No missing data problem for ICE subjects — every subject has an outcome. Straightforward to implement, especially for binary endpoints.
  - **Limitation**: Results can be misleading if the ICE does not truly reflect treatment failure. For example, if patients in the control arm are disproportionately rescued for reasons unrelated to efficacy, the strategy may favor the treatment arm unfairly. Defining a clinically meaningful "bad score" for continuous endpoints requires careful team discussion.

Intercurrent Strategies

- **While-on-Treatment Strategy**
  - **Objective**: Evaluate the treatment effect using only data collected prior to the occurrence of the ICE — the question of interest is limited to the period when the patient was on their assigned treatment.
  - **How it works**: Data collected after the ICE are ignored; only pre-ICE observations contribute to the analysis. The endpoint is effectively evaluated at the last observed time point before the event, rather than at a fixed landmark.
  - **Advantage**: No imputation needed for post-ICE data. Straightforward to implement when pre-ICE data are well-collected.
  - **Limitation**: If patients discontinue at very different time points across arms, the strategy compares outcomes measured at different stages of follow-up — making results difficult to interpret. A drug with strong early activity but diminishing long-term benefit could be wrongly favored. This strategy is generally not appropriate when the ICE itself signals that the treatment is no longer working.

- **Principal Stratum Strategy**
  - **Objective:** Estimate the treatment effect within a specific subgroup — the principal stratum — defined by patients' potential ICE status under both treatment conditions (e.g., adherers, patients who would not use rescue medication regardless of which arm they were assigned to).
  - **How it works**: Patients are grouped into latent subgroups based on their ICE behavior under each treatment. Because stratum membership is partially unobserved (you cannot know what would have happened under the alternative treatment), it must be inferred through statistical modeling and assumptions.
  - **Advantage**: Preserves causal validity by defining subgroups based on pre-randomization potential outcomes rather than post-baseline events. This allows researchers to isolate the true therapeutic effect of the treatment in a specific cohort without breaking randomization or introducing the selection bias inherent in traditional sub-group or per-protocol analyses.
  - **Limitation**: The principal stratum is *latent* and cannot be directly observed, requiring untestable assumptions that can meaningfully affect results. The identified subgroup may not represent the

Intercurrent Strategies

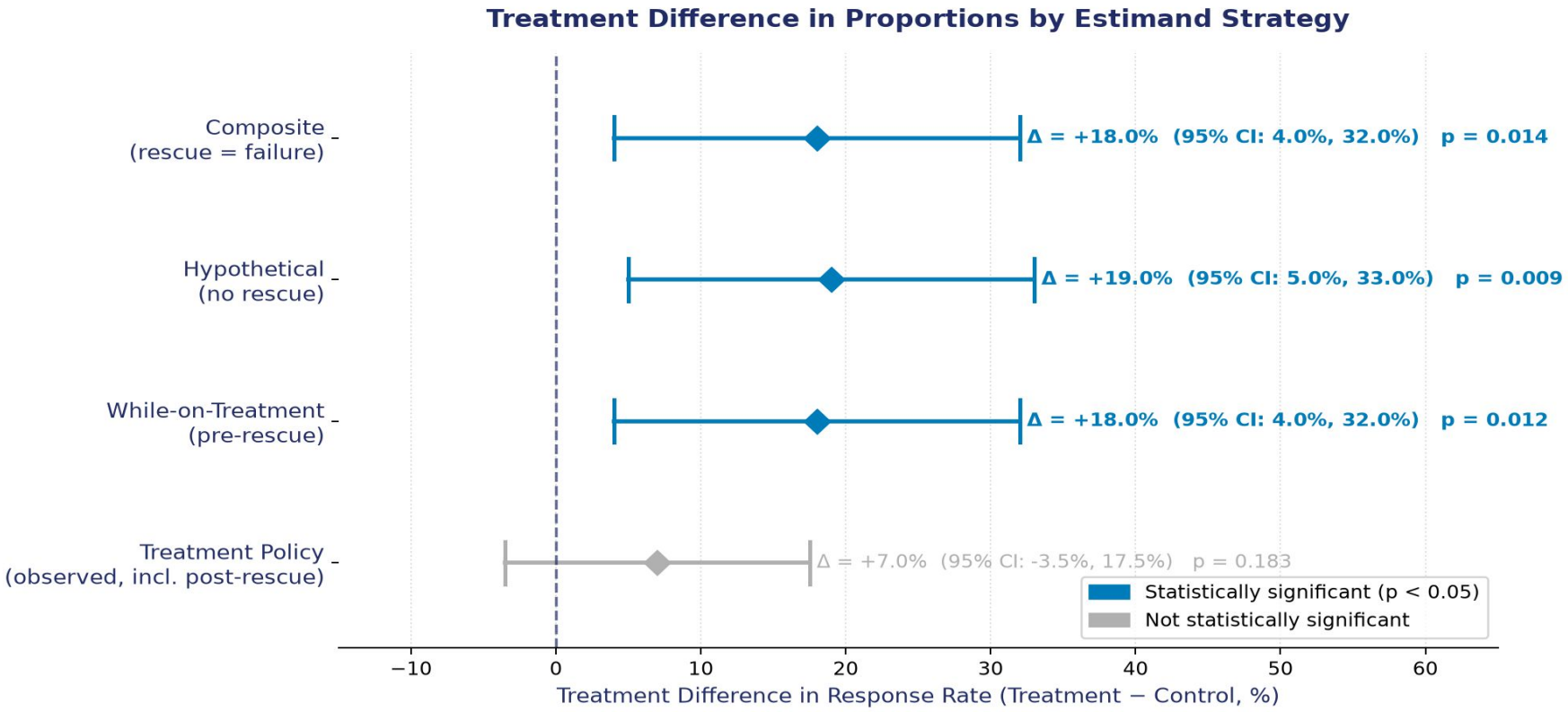


- **Patient 1** represents subjects who completed the study without treatment discontinuation or a need for rescue medication and have efficacy outcome at all time points.
- **Patient 2** represents subjects who completed the study and provided efficacy outcome at all months. However, because a rescue medication is used at Month 4, the efficacy outcomes at the subsequent months are affected by the rescue medication.
- **Patient 3** represents subjects who received a rescue medication and discontinued the study thereafter. Consequently, an efficacy outcome at Month 8 is missing.
- **Patient 4** represents subjects who discontinued the study treatment due to toxicity/lack of efficacy but remained in the study and provided efficacy data at subsequent months.
- **Patient 5** represents subjects who discontinued the study treatment due to lack of efficacy/toxicity and left the study without providing efficacy data at subsequent months.
- **Patient 6** represents subjects who are lost to follow up and therefore, efficacy data is missing for all visits after Month 2.

Example: Hypothetical Case Study

- Design: A 52-week randomized double-blind placebo-controlled study.
- Randomization: 200 subjects were randomized in a 1:1 ratio to the investigational arm or placebo.
- Primary endpoint: Proportion of subjects with treatment success at Month 52.
- Key ICE: Escape therapy.
- Population level summary: Difference in proportion of success at Month 52.
- Analysis: Difference and 95% CI were derived adjusting for stratifying variables.

Example Case Study: Results Under Different Strategies



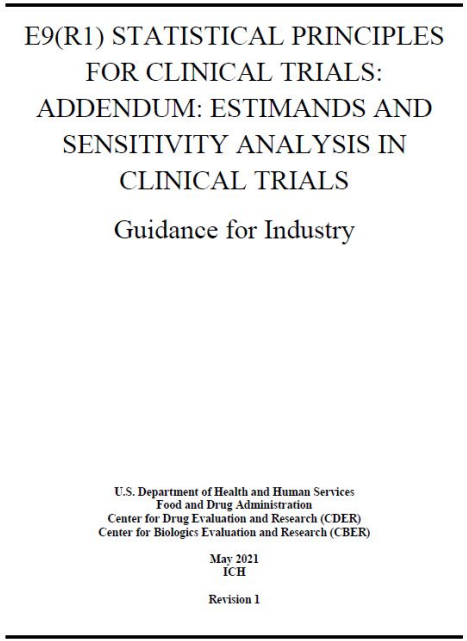
Key Take Aways

- The estimand needs to be defined prior to the design of the study
- Statisticians alone cannot define the estimand — it encodes clinical judgment.
- Clinicians must answer:
  - What is the clinical question of interest?
  - Which post-randomization events should be classified as ICEs?
  - Is it clinically meaningful to observe patients after treatment discontinuation?
  - Does an ICE like rescue medication use represent treatment failure — or routine care?
- Statisticians then translate those answers into a precise estimand and choose the analysis method that targets it.
- The estimand is NOT driven by what is statistically convenient — it is driven by what is clinically meaningful.

Part 2: Missing Data

- Defining missing data and its impacts
- Prevention of missing data
- Types of missing data
- Missing data and analyses
- Sensitivity analyses

What is “missing data”?


E9(R1) STATISTICAL PRINCIPLES FOR CLINICAL TRIALS: ADDENDUM: ESTIMANDS AND SENSITIVITY ANALYSIS IN CLINICAL TRIALS

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2021
ICH

Revision 1

https://www.fda.gov/media/148473/download

*Data that **would be meaningful** for the analysis of a given estimand but **were not collected**. They should be distinguished from data that do not exist or data that are not considered meaningful because of an intercurrent event.*

Note: Choice of estimand determines what missing data means!

What is “missing data”?

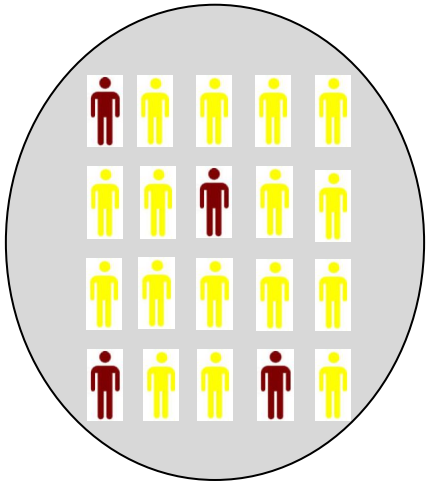
- Consider a 104-week study with endpoint evaluated at Week 52:

- Subject | Event | ICE? | Missing Data?
  - 01 | Died before Week 52 | Yes | No (data cannot exist after death)
  - 02 | Used prohibited medication | Yes | Depends on estimand strategy
  - 03 | Used prohibited med + completed study | Yes | No (Week 52 data observed)
  - 04 | No ICE, no Week 52 data collected | No | Yes — classic missing data
  - 05 | Discontinued treatment, left study | Yes | Yes (if treatment policy used)

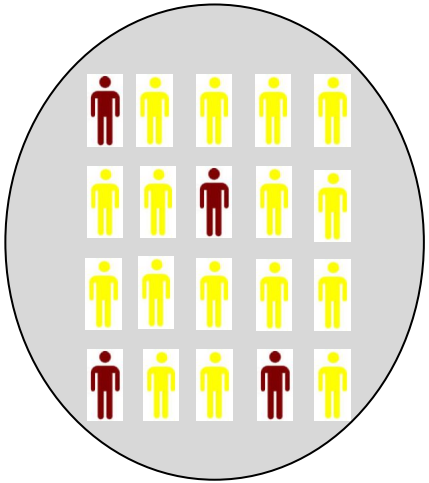
- The same subject can be an ICE case OR a missing data case depending on the estimand — which is why defining the estimand first is critical.

Example: Complete Data

Consider the following “complete data” (i.e., data that would be meaningful for the analysis of our estimand).



Treatment



Control



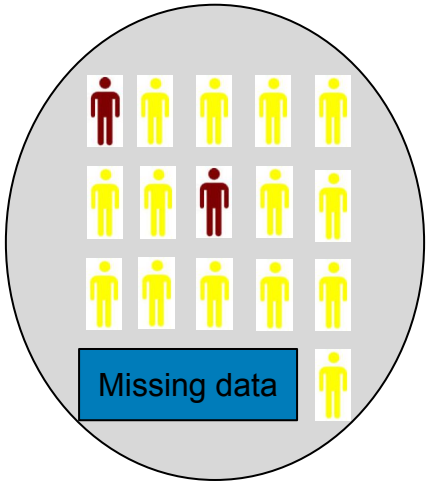
Analyzing the complete data

$\Longrightarrow$ Observed event rate of 20% on both arms

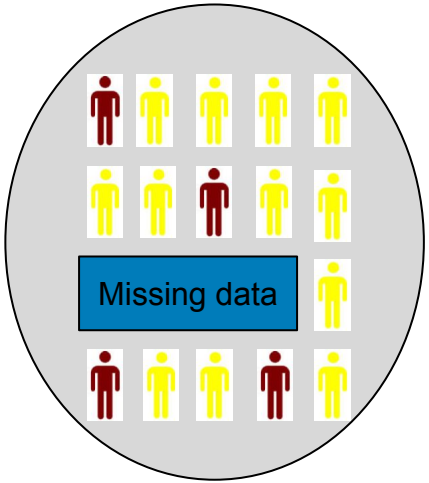
Rate Ratio of 1, observed treatment effect indicates no difference in risk between groups

Example: Missing Data

Now, consider if some information is missing on each treatment arm



Treatment



Control

Event No event

If we were to analyze the remaining data as if it is complete

$\Rightarrow$ Observed event rates of 12.5% (2/16) and 25% (4/16)

Rate Ratio= 2, observed treatment effect indicates a 2-fold decrease with treatment

Impacts of Missing Data

- Reduction in efficiency (statistical power)
  - Loss of information
  - Increased variability
- May bias estimation, depending on the type of missingness
- Reduce representativeness of samples
  - Results may not be generalizable outside of the study
- Complicate analyses!

Prevention of Missing Data

- Easy solution to handling missing data, on the surface, is to not have any (i.e., prevent it!)
- The **study design** should be consistent with minimizing attrition, and **study conduct** should include efforts to retrieve **meaningful data (ties to estimand)**
- Requires careful examination and consideration at **the protocol review stage**

Treatment Discontinuation

- Consider an estimand that uses the treatment policy strategy for treatment discontinuation:
  - The value of the endpoint is relevant even after a patient discontinues treatment
  - If data are not observed after discontinuing treatment, this results in missingness
- However, we need to distinguish between treatment discontinuation and study withdrawal

Differentiate Treatment and Study Withdrawal

- Statistical analysis plans and informed consent forms should clearly differentiate treatment discontinuation from study withdrawal
- Possible reasons for stopping treatment
  - Lack of efficacy, adverse event, lab abnormality, physician/investigator decision, tolerability, etc.
  - In this case, continue to follow patients for all regularly scheduled efficacy and safety
- Possible reasons for withdrawing from study early
  - Withdrawal of consent or successful ascertainment of all safety and efficacy information or loss to follow-up

Avoid Inadvertently Encouraging Attrition

- Many statistical analysis plans contain language instructing clinical centers to withdraw patients in violation of the protocol
  - For non-compliance with study medications, missed visits, etc.
  - Example protocol text: *"If non-compliance during the treatment period continues, the subject may be terminated from the study at the investigator's discretion."*
- If patient's safety is assured, non-compliance does not require early withdrawal

Planning for Analysis

- Pay attention to:
  - The estimand of interest
  - Whether the analysis will provide a reliable and interpretable estimate of treatment effect and appropriate hypothesis testing/confidence intervals
  - The assumptions required to implement the analysis
    - Including missing data assumptions
    - Ideally, prospectively documented and justified
    - Implausible assumptions avoided

Note: Usually an iterative process should be used to reach an estimand that is of **clinical relevance for decision-making**, and **for which a reliable estimate** can be made (generally includes reasonable assumptions)

Types of Missing Data

- In considering an analysis, statisticians often refer to three “mechanisms” of missing data:
  - Missing completely at random (MCAR)
  - Missing at random (MAR)
  - Missing not at random (MNAR)
- Used to describe different types of missing data and some of the assumptions that underpin statistical approaches to handling missing data.
- Helps us consider situations when some methods will work to provide valid inference and others will not.

Rubin DB. Inference and missing data. *Biometrika.* 1976;63:581–592.

Missing Completely at Random

- Probability that the data are missing is not related to either the specific value which is supposed to be obtained or the set of observed responses
- No systematic differences exist between participants with missing data and those with complete data
- Assumes patients with missing data are a random sample of randomized patients
- Examples:
  - Spirometer measuring pulmonary function randomly fails
  - A weighing scale ran out of batteries. Some of the data will be missing simply because of bad luck

Missing At Random

- Probability that responses are missing depends only on observed study variables (baseline characteristics and/or earlier outcomes)
- Data are missing systematically related to the observed but not the unobserved data
- Example:
  - Dropout is more likely for men compared to women, but all men have the same chance of dropping out and all women have the same chance of dropping out
  - Subjects who are discharged from the hospital early in the study are more likely to have missing data at the end of the study than those who are not discharged (probability of missingness depends on discharge status)

Missing Not at Random

- If not MCAR or MAR, then MNAR
- Sometimes called “non-ignorable”
- Probability of missingness depends on values of unobserved variables, including the missing data themselves
- Example:
  - Patients with worsening outcomes more likely to be lost to follow-up (and not observed for these outcomes)
  - Participants in a public opinion poll with weaker opinions respond less often (and this information cannot be observed)

Some Common Missing Data Handling Techniques in Analysis of Clinical Trials

- Complete Case Analysis (also “listwise deletion”)
- Single Imputation Approaches
- Multiple Imputation Approaches
- Other Approaches (Mixed Models, Generalized Estimating Equations)

Complete Case Analysis

- Ignores missing data and performs the analysis with complete data only
- Very easy to implement, simple
- If the assumption of MCAR is satisfied, a complete case analysis is known to produce unbiased estimates
  - BUT can still be a loss of efficiency (data are discarded)
- When not MCAR (i.e., MAR or MNAR), this approach may result in biased estimates
  - Recall our first example

→ Generally, not recommended for pivotal trials

Single Imputation Methods

- Replace a missing data point by a single value and analyses are conducted as if all the data were observed.
- Easily implemented
- Single imputation examples:
  - Last observation carried forward (LOCF)
  - Baseline observation carried forward (BOCF)
  - Worst case imputation
- Often perceived to provide conservative estimates but not universally true
- In many cases do not align with **scientifically plausible assumptions** about the mechanism that produced the missing data
- **Does not fully capture the uncertainty** that the imputed value is correct. Thus, these methods underestimate the variability of treatment effects, can lead to inappropriately low p-values (draw incorrect conclusions).

Multiple Imputation Methods

- Multiple imputation methods consist of three phases: imputation, analysis, and pooling.
  - Imputation: We first generate multiple copies of the original dataset by replacing missing values using a model based on observed data (imputation model), which leads to multiple imputed datasets
  - Analysis: Analyze each imputed dataset as complete set and generate the parameter estimates of interest
  - Pooling: Combine the parameter estimates across the datasets to produce a final point estimate and standard error
    - Takes into account the uncertainty of the imputation process
    - In contrast to single imputation, risk of underestimating the variance is reduced
- Can produce unbiased estimates and correct p-values, but still relies on assumptions being appropriate (e.g., MAR, construction of the imputation model and data used for imputation)

Other Missing Data Approaches

- Certain statistical methods do not use explicit imputation
- For example, mixed models (e.g., MMRM) can be used to handle missingness when outcomes measured repeatedly over time
- Likelihood methods find estimates of the treatment differences that maximize the probability of the observed data
- Can produce unbiased estimates and correct p-values but again **relies on assumptions** about missingness mechanism + distributional assumptions (e.g., normality) in the modeling

Caution with MAR

- Standard statistical analyses (e.g. MMRM, Cox proportional hazards) are reliable only if **missing data are missing at random (MAR)** given model
- Assumption often questionable because patients who drop out tend to be systematically different (e.g., are no longer on treatment)
- Nothing in your data can tell you whether missing data are missing at random
  - Always evaluate potential impact of missing not at random data on conclusions (**sensitivity analyses**)

Sensitivity Analyses

- ICH E9(R1) definition: A series of analyses conducted with the intent to **explore the robustness** of inferences from the main estimator [from the main analysis] to deviations from its underlying modeling assumptions and limitations in the data
  - Evaluate whether conclusions hold up under plausible violations in analysis assumptions – i.e., what if our assumptions are wrong?
  - Note: we will focus on missing data assumptions, but sensitivity analyses can also be designed to assess other assumptions (e.g., distributional)

Sensitivity vs Supplementary Analyses

- Note that sensitivity analyses are distinct from supplementary analyses, though both are useful tools.
  - Key Supplementary analyses ask: *What else can we learn about the treatment effect?*

Supplemental analysis

- **Examples of Supplemental Analyses (Estimands & Missing Data Context)**
- **Estimand-focused supplemental analyses:**
  - **Treatment policy vs. hypothetical strategy** — The main analysis may use a treatment policy estimand (outcomes regardless of treatment discontinuation), while a supplemental analysis uses a hypothetical estimand (e.g., *"what if all patients had remained on treatment?"*), offering a different lens on the underlying treatment effect
  - **While-on-treatment estimand** — Restricting the analysis window to the period patients were actually on treatment, as a complement to an ITT-based primary estimand that includes post-discontinuation outcomes
  - **Principal stratum estimand** — Evaluating treatment effect within a subgroup defined by post-randomization behavior (e.g., patients who would adhere to treatment regardless of assignment), exploring a distinct causal question from the primary analysis

Choosing Sensitivity Analyses

- Start with considering and acknowledging the assumptions of the main analyses (ideally done when specifying main analysis)
- Structure sensitivity analyses to assess these assumptions in an understandable and systematic way
  - Note: Altering many aspects of main analysis simultaneously may lead to challenges in identifying which assumptions, if any, are responsible for potential differences
- Pre-specification ideal but sensitivity analyses are important regardless
- Do not restrict to single or limited number of alternative assumptions, should be series of analyses to ensure conclusions hold up across plausible assumptions
  - Ideally, a priori discussions about plausibility of assumptions

How much missingness requires sensitivity analyses?

- Degree of sensitivity to violations in missing data assumptions depends on many factors
  - Very difficult to predict
  - There is no minimal amount of missing data below which missing data sensitivity analysis is not important

Tipping Point Analyses

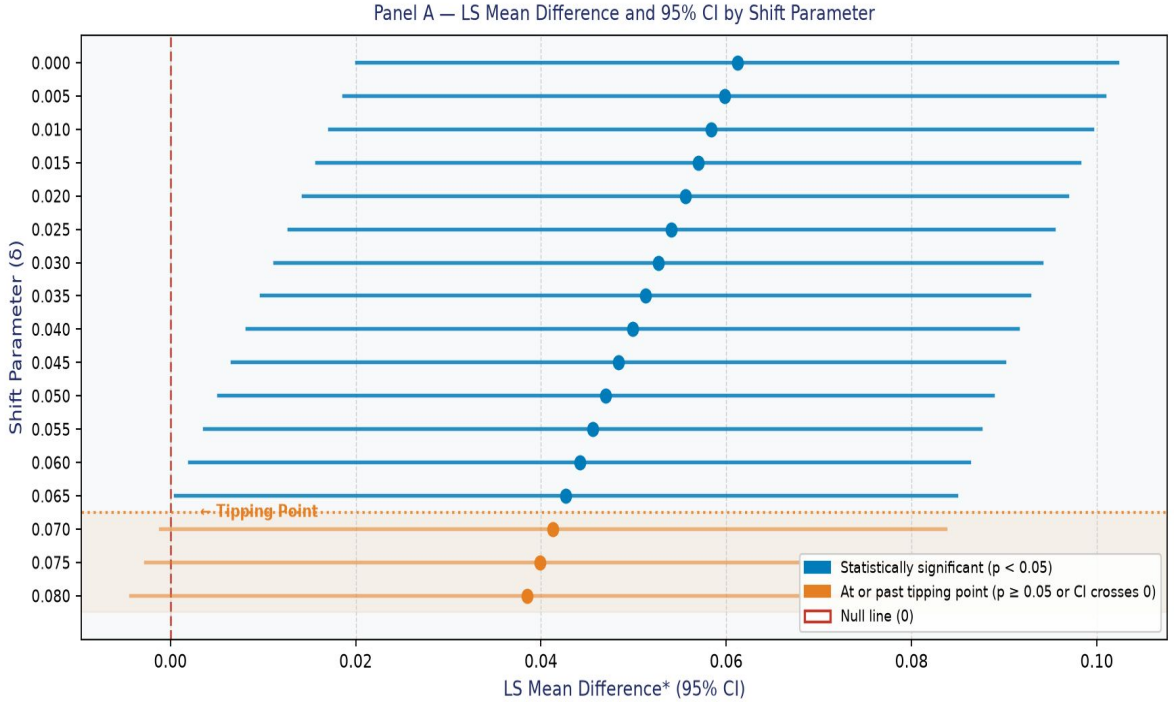
- One type of sensitivity analysis for assessing missing data
- Treatment effect is re-evaluated after systematically and comprehensively varying assumptions about the missing outcomes on the two treatment arms
- Goal is to identify the “tipping point” – point where the conclusions (results) change, and facilitate cross-disciplinary discussions about the plausibility of such assumptions
- Note:
  - Preferably, should be two-dimensional, i.e., allow assumptions about the missing outcomes on the two arms to vary independently
  - Include scenarios where missing data on drug tend to have worse outcomes than missing data on control

Tipping Point Analysis (Example)

- Estimand: Difference in the mean change from baseline at Month 24 between Active and Control regardless of any non-adherence to or interruption of study treatment, and regardless of initiation of alternative treatment (Treatment policy).
- Main analysis: Linear mixed effects model with visit, treatment, treatment by visit interaction and the stratifying variables included.
- Assumption: Missing data at Month 24 are **MAR** conditional on model factors.
- Missing Data: Amount of missing data: Roughly 5–10%.
- Tipping point: Shift parameters with increments of 0.05 were added to the imputed value.

Example Tipping Point Analysis

- Key finding: The tipping point was reached when a shift parameter of 0.065 was added to the imputed values for the treatment arm.
- Whether this reflects robustness of the results depends on the outcome.



*LS: Least squares estimate

Summary and Final Thoughts

- Missing data are problematic
  - Impact the reliability of estimates in terms of bias and variability
- Best approach is prevention
  - Carefully prespecify what trial intends to estimate
  - Ascertain the primary endpoint on all randomized patients who do not withdraw consent to be followed
- Plan and pre-specify how missing data will be addressed in analysis
  - Minimal, reasonable, clearly stated assumptions
  - Analyses to explore sensitivity of conclusions to violations should be pre-specified, conducted, reported

Backup Slides