IMPORTANCE OF CENSORING MECHANISMS IN SELECTING APPROPRIATE ESTIMANDS

JONATHAN SIEGEL (BAYER)
MICHELLE CASEY (PFIZER)
HANS-JOCHEN WEBER (NOVARTIS)
ANJA SCHIEL (EMA BSWP/SAWP)
STEFAN ENGLERT (ABBVIE)
STEVEN SUN (JOHNSON & JOHNSON)
KASPAR RUFIBACH (HOFFEMANN-LA ROCHE)

DRAFT

July 10, 2019

DISCLAIMER

- This presentation does not represent the views of any of the co-authors' employers, or any regulatory agency.

- This is a draft of the presentation and has not received final approval for release.
  A final version of the slides will be presented at Joint Statistical Meetings, Denver, CO (USA) on July 31, 2019

THE ESTIMANDS FRAMEWORK: KEY CONCEPTS

- Draft addendum to ICH E9 guidance on statistical practice in clinical trials for drug approval

- An **estimand** attempts to define prospectively and contextualize the outcomes of a treatment observed in a specific study setting.

  - It defines s population-level noumenon that an estimator attempts to estimate and infer to based on phenomena observed in a trial.

  - By exploring different estimands, an assessment of what would have happened hypothetically under different treatment conditions is possible.

  - It requires defining a population of inference, a variable or endpoint, a specification of how to account for intercurrent events, and a population-level summary (statistic) serving as the basis for comparison

- **Intercurrent events** occur after treatment initiation and either affect interpretation of the variable or preclude its observation

- It is important to clearly distinguish between intercurrent events and missing data.

  - Missing data will weaken the strength of the estimator without altering or biasing the estimand

  - Intercurrent events alter the estimand and require a distinct statistical approach to allow their interpretation.

A PARADIGM SHIFT

- Traditionally, statistical methods have often generally assumed that missing data is non-informative or missing at random

- Clinical trialists have challenged these assumptions as often being unrealistic in clinical trial situations

- The estimands framework represents a response to this call for rethinking. It focuses on:
  - The concept of intercurrent events which alter or bias the estimands as distinct from missing data which does not
  - Strategies to avoid or reduce missing data and intercurrent events where possible, and collect events of interest under any circumstances and preferably beyond the primary outcome
  - Strategies to address intercurrent events
    - Assumptions behind and appropriate use of intercurrent event strategies
    - Upfront articulation of assumptions and post-trial sensitivity analyses to check them

In an estimands framework, it is necessary to understand the actual reasons for intercurrent events, understanding the impact these events might have on the interpretation of the actual data in light of the research question to be answered and pre-plan for them in close cooperation among study team members of different disciplines.

IMPACT ON ONCOLOGY

- Oncology development tends to use time-to-event variables such as overall survival and progression-free survival

  - The very length of trials therefore can influence the possibility of intercurrent events to occur.

- Oncology drugs often have complex safety profiles

  - Decisions to withdraw from treatment are often related to perceived treatment efficacy and safety

  - High-mortality disease increases ethical imperative to move patient to a new therapy/trial if current therapy appears non-beneficial

- Standard assumptions often require following patients past end of treatment or into a new trial

- Oncology trials are often particularly dependent on non-informativity assumptions, in clinical trial settings where these assumptions may be particularly untenable

- Most patient- and physician-initiated reasons for discontinuation have at least potential for informativeness

MISSING DATA, TYPES OF INTERCURRENT EVENTS, AND STRATEGIES

| Missing Data	**Positively informative** provide qualitative information about the event of interest	**Counterfactual** confound the event of interest	Irrelevant
• Does not introduce bias or alter the estimand. • Censoring assumes this. **Primary Strategy:** • **Treatment Policy**	• Scientific question is what actually happened, including the intercurrent event • Goal of improvement is to better incorporate the intercurrent event into the analysis **Primary Strategy:** • **Composite**	• Scientific question is what would have happened if intercurrent event had not occurred. • Intercurrent events rendered uninformative conditioned on a model **Primary Strategies:**	• Scientific question is about what happened prior to the intercurrent event **Primary Strategy** • **While on Treatment**

No way to determine informative censoring definitely

Note: The strategies presented in this overview slide will be discussed in more detail in this presentation
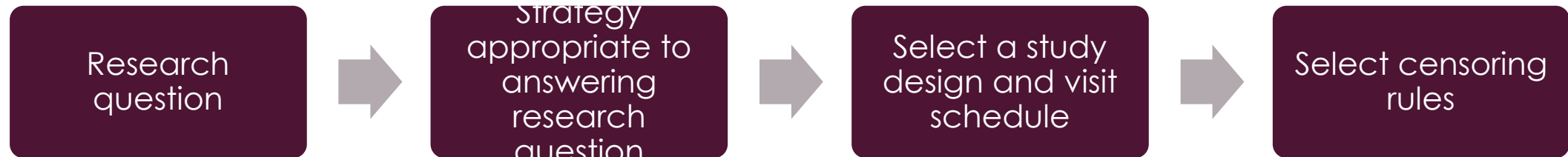
STRATEGY SELECTION AND FEASIBILITY

- A clear scientific objective will help in selecting estimands and strategies

- Team needs to assess the clinical relevance and feasibility of estimating the desired estimand in the proposed setting.

- Should the desired approach have feasibility issues, then alternative estimands and strategies should be considered

- There is often no perfect strategy. The initial or conventional strategy, even with feasibility issues, might still be better than the alternatives.
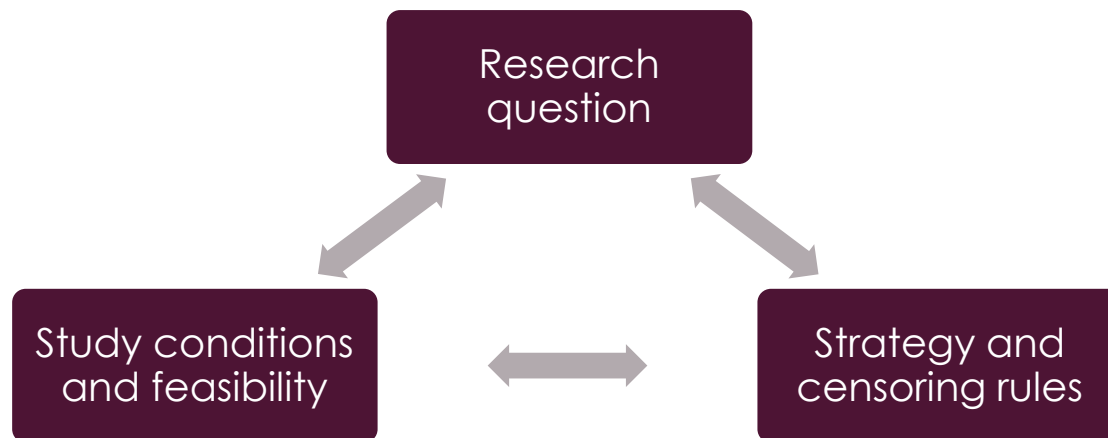
Discuss the best estimand

Assess Scientific Objective, Trial Design & Feasibility

Explore alternative Strategies
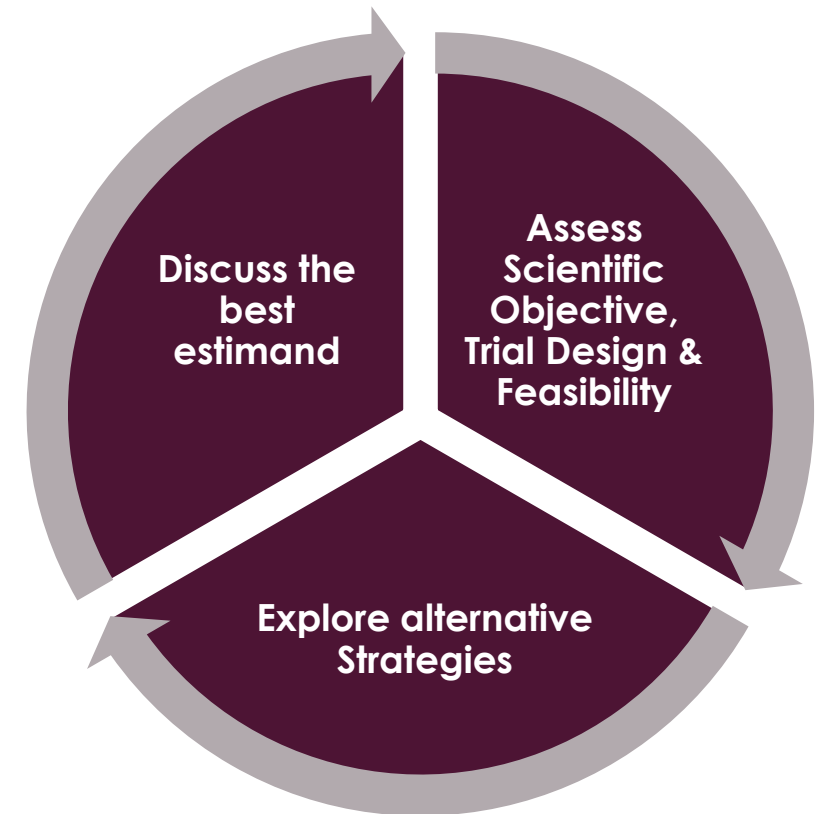
STUDY DESIGN MAY BECOME A LESS LINEAR PROCESS

- As traditionally conceived, study design is a linear process:

Research question	→	Strategy appropriate to answering research question	→	Select a study design and visit schedule	→	Select censoring rules

- It may need to become more non-linear in practice.

Research question

Study conditions and feasibility

Strategy and censoring rules

The procedure needed to address this non-linearity may more closely resemble W. Edward's Deming's "Plan-Do-Study-Act" cycle (Deming, 1986) than a classic linear "waterfall" procedure.

Feasibility may influence research questions

BETTER TACTICS: COLLECTING DATA

- The estimands framework affects how we select strategies and design studies

- But it also depends on new and better approaches to very nuts-and-bolts issues like data collection.

- Comprehensively identifying and accurately classifying potential intercurrent events is critical.

- We need to know
    - Why patients did not come to clinic
    - Why patients withdrew from treatment and/or study follow-up
    - Reasons for and informativeness potential of other on-study events that may influence the results.

- Improved data collection is key to making the framework workable in practice
    - Identify and collect data on all potential intercurrent events
    - Follow up and document why data is missing

- Existing data categories need to be revisited and optimized to new estimands purpose

STRATEGIES: "ITT"

- We will present each of the major estimands strategies of special relevance to oncology with appropriate examples.

- We'll begin with what is popularly known as the "ITT" approach, a work-horse of current clinical trials, and discuss how the estimands framework will affect it.

  - Fleming et al. (2009) advocated following patients until the event of interest is document, and censoring all patients without it, regardless of intervening events

  - This approach commonly known as the "ITT" (Intent-to-treat) approach became incorporated in the 2018 FDA and 2017 EMA guidance on PFS for blinded, randomized trials.

- It mostly corresponds to a **treatment policy strategy** in the estimands framework.

  - In a treatment policy strategy, patients are followed past intercurrent events, and intercurrent events are assumed to be uninformative and ignored.

"ITT" ISSUES

- In the estimands framework, a treatment policy strategy requires a study design and follow-up program that meets the underlying assumptions

  - Assessing patients on a consistent, treatment-independent visit schedule.

  - Following patients until and beyond the observation of outcome of interest.

- This follow-up may sometimes not be feasible in practice.

- Feasibility should be evaluated, depending on study context, and alternative strategies should be considered.

  - Feasibility of follow-up beyond end of treatment may depend on practical factors including availability of other clinical trials, frequency and intrusiveness of assessments compared to benefit of continued care, distance from clinic, etc.

- Even when the necessary follow-up is not feasible and some informative censoring and resulting bias is anticipated, a treatment policy strategy might still be the best available strategy.

HYPOTHETICAL STRATEGY FOR ADDRESSING COUNTERFACTUAL INTERCURRENT EVENTS

- A hypothetical strategy asks what would have happened if the intercurrent event hadn't occurred.

- Censoring for subsequent therapy or other intercurrent events, a standard approach in past FDA guidance, can be interpreted as a hypothetical strategy

- It describes what would have happened if:

  - The patient had not experienced subsequent therapy or other intercurrent event, and

  - The patient's hazard of the event of interest was the same after the intercurrent event as in patients who did not experience it.

- Because patients may change treatments due to unsatisfactory efficacy, these assumptions may be problematic.

- In addition, if treatment switching is a common phenomenon in the study, censoring large numbers of patients reduces study power, sometimes drastically.

HYPOTHETICAL STRATEGY: CONSIDERATIONS & EXAMPLES

- Causal inference methods to address treatment switching (crossover) represent another hypothetical strategy example
  - Examples include rank-preserving structural failure time (RPSFT) and inverse probability weighting (IPW)
  - These methods require strong assumptions
    - Causal methods assume all systematic information about treatment effect is captured in the variables modeled
    - Cancer is a poorly understood disease, and we often don't even know enough to predict which patients will benefit from treatment
- The estimands framework provide conceptual tools to frame the relevant questions and to evaluate hypothetical strategies for potential use in an appropriate setting
  - Explicit mention of hypothetical strategies
  - The ability to frame custom research questions specific to a development program
  - The ability to justify assumptions, where they do approximately hold, using post-hoc sensitivity analyses.

WHILE ON TREATMENT STRATEGY: CONSIDERATIONS & EXAMPLES

- The "while on treatment strategy" poses a research question that is only interested in the treatment effect until the intercurrent event occurs.

- A classic example is a purely palliative treatment.

  - Treatment purpose is to make the patient comfortable, and not to alter survival.

    - The effect of interest is improvement or worsening of symptoms prior to death.

    - Death does not represent a negative outcome for a palliative treatment TTE analysis.

  - To reflect the interest in only assessing palliative benefit until death, death is more appropriately modeled as a competing risk event than a censoring event.

    - Death is not a negative outcome (not an event), but takes the patient out of the risk set.

WHILE ON TREATMENT STRATEGY: CONSIDERATIONS & EXAMPLES

- Time to progression (TTP), a familiar although no longer commonly used endpoint, represents a similar example in an oncology setting
  - Modeled using Kaplan-Meier, as has been standard, TTP can be interpreted as a hypothetical strategy
    - We ask what would have happened if patients hadn't died
    - We assume patients with deaths would have had risks similar to surviving patients during the time period following their death (interpreting non-informative censoring in this context)
  - A while-on-treatment strategy asks a different question, uses a different method, and makes a different estimate
    - We ask what is the risk of progression while the patient is alive
    - The competing risks approach applies this interpretation by removing patients with deaths from the risk set without counting them as events, generating an estimate which is unbiased for this specific question.
- While-on-treatment strategies and associated competing risk approaches can also come up in other areas in oncology, including time to cause-specific events and time to non-fatal safety events.

COMPOSITE STRATEGY FOR ADDRESSING POSITIVELY INFORMATIVE INTERCURRENT EVENTS

- Under an estimands framework, a composite strategy may be appropriate when intercurrent events are highly correlated with or highly related to the outcome of interest.

- In a TTE context, a new endpoint is created which is the earlier of the event of interest and the intercurrent event

  - Progression-free survival, a composition of time to progression and overall survival, is a well-known oncology example

- When assessing PFS, if patients are particularly likely to leave the study due to clinical progression without waiting for formally documented progression, a composite strategy including clinical progression might be appropriate to consider

PRINCIPAL STRATUM STRATEGY: CONSIDERATIONS

- The principal stratum strategy (Frangakis and Rubin, 2002) typically attempts to define the population of interest as the patients in whom the relevant intercurrent event is expected not to occur

- Patients are classified into four strata:

    - 1.those who do not experience the intercurrent event regardless of which treatment they were assigned

    - 2.those who would always experience the intercurrent event regardless of which treatment they were assigned

    - 3.those who only experience the intercurrent event if assigned to the treatment group

    - 4.those who only experience the intercurrent event if assigned to the control group

- Classification is modeled based on characteristics known at baseline.

- Generally outside scope of censoring mechanisms framework.

REVIEW: MISSING DATA, TYPES OF INTERCURRENT EVENTS, AND STRATEGIES

Missing Data	Positively informative provide qualitative information about the event of interest	Counterfactual confound the event of interest	Irrelevant
• Does not introduce bias or alter the estimand. • Censoring assumes this. **Primary Strategy:** • **Treatment Policy**	• Scientific question is what actually happened, including the intercurrent event • Goal of improvement is to better incorporate the intercurrent event into the analysis **Primary Strategy:** • **Composite**	• Scientific question is what would have happened if intercurrent event had not occurred. • Intercurrent events rendered uninformative conditioned on a model **Primary Strategies:**	• Scientific question is about what happened prior to the intercurrent event **Primary Strategy** • **While on Treatment**

No way to determine informative censoring definitely

EVALUATING AND PROPOSING ALTERNATIVE STRATEGIES

- The guidance emphasizes the importance of discussing strategies with regulatory authorities in advance

- In order to obtain acceptance of a strategy, a sponsor needs to show:
  - The strategy is appropriate to address the scientific objective
  - The assumptions required for the strategy are reasonable in the study context
  - Alternatively, applying the strategy will result in less bias, taking into account the research question and study context, than accepted methods such as the standard treatment strategy

- It might sometimes be impossible to show that a proposed strategy is a fully appropriate one (unbiased, all assumptions hold), but nonetheless possible to show that it is less biased, or otherwise less inappropriate strategy than a more conventional strategy.

- Practical considerations including assessments of likely patient behavior may have significant impact on decisions regarding strategy

EXAMPLE: SEQUENCE OF TREATMENTS

- In the Bayer 16507 study, (Sternberg et al., 2018), 6 doses of Xofigo (Ra-223) at standard dose were compared with either 12 doses at standard dose or 6 doses at a higher dose, all q4w.

- For the 6 dose/12 dose comparison, both regimens are the same for the first 6 doses (~24 weeks)
  - Defining estimand from first treatment would result in a delayed effect for first 6 doses
    - Delayed effects reduce study power and obscure any treatment benefits

- Approach used was to redefine both the **population of interest** and the **treatment effect** in a manner designed to facilitate estimation
  - Define the population of interest as the patients who are eligible to start further treatment at the change point
  - Define the treatment effect as the effect of treatment after the change point
  - Define the start of TTE endpoints as the time of the therapy change point.

EXAMPLE: SEQUENCE OF TREATMENTS

- By redefining the population of interest, the treatment effect, and the start point of TTE variables:
    - The complications of modeling changes in hazards resulting from change in therapy are avoided
    - The ability to retain standard proportional hazards assumptions is preserved.
- Similar issues arise in other contexts, such as CAR-T therapies (Neelapu et al, 2017; Schuster et al., 2019),
- In some cases the research question may be redefined into one that may not be of primary interest to a treating physician or patient
    - They may be interested in assessing the complete effect of the entire regimen
    - By starting measuring the treatment effect in the middle of the regimen, this purpose is frustrated.
    - Unlike a 12-dose regimen which can reasonably be divided into an independently beneficial 6-dose induction regimen followed by a 6-dose maintenance regimen, separating the CAR-T manufacturing period from the study observation clock may be a more questionable division.
- Feasibility considerations may sometimes suggest designing a study to answer a question that can be answered reliably under the circumstances, not the one we really want to answer.
- It may also be appropriate to have different estimands of the "same" endpoint for different research purposes and audiences.

OPENING THE BOX

- These examples illustrate how different estimands and corresponding estimators correspond to different research questions, analysis methods, censoring rules, and interpretations for what was previously regarded as a single endpoint with a single standard estimation method, censoring procedure, and meaning.

- The estimands framework potentially provides additional opportunities to customize endpoints to specific research questions and contexts.

- Careful identification of which intercurrent events are appropriately classified as endpoint events, which as censoring, and which as competing risk is critical.

- Careful reframing of the research question may also yield an appropriate way to address problematic intercurrent events.

SUMMARY & CONCLUSIONS

- The estimands framework is intended to result in statistical assumptions that better reflect the realities of clinical trials

    - It emphasizes appropriate strategies to handle intercurrent events, based on realistic and justified assumptions about their informativeness

    - It is to some degree a categorization and reframing of existing methods, but also opens the door, at least potentially, to new strategies and possibilities

- Understanding censoring mechanisms is critical to appropriate use of the estimands framework in a time-to-event context

- Greater cooperation between statistical methodologists, applied statisticians, clinicians, and trial operational specialists is critical to success.

QUESTIONS?

THANK YOU!

JONATHAN SIEGEL (BAYER)
MICHELLE CASEY (PFIZER)
HANS-JOCHEN WEBER (NOVARTIS)
ANJA SCHIEL (BSWP/SAWP)
STEFAN ENGLERT (ABBVIE)
STEVEN SUN (JOHNSON & JOHNSON)
KASPAR RUFIBACH (HOFFEMANN-LA ROCHE)

REFERENCES

- European Medicines Agency, Committee for Human Medicinal Products. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. Draft Step 2b. (2017)

- Deming, WE. *Out of the Crisis*. MIT Press (1982)

- Fleming, T, et al. Issues in using progression-free survival when evaluating oncology products. *J Clin Oncol* 27:2874-2880 (2009)

- Frangakis, CE and Rubin, DB. Principal stratification in causal inference. *Biometrics* 58:21-29 (2002)

- Sternberg, et al. A randomized phase 2 study investigating 3 dosing regimens of radium-223 dichloride (Ra-223) in bone metastatic castration-resistant prostate cancer (mCRPC). *J Clin Oncol* 36 (suppl; ASCO abstr 5008) (2018)

- Neelapu, SS et al. Axicabtagene ciloleucel CAR T-cell therapy in refractory large B-cell lymphoma. *N Engl J Med.* 377:2531-2544 (2017)

- Schuster, SJ et al. Tisagenlecleucel in adult relapsed or refractory diffuse large B-cell lymphoma *N Engl J Med* 380:45-56 (2019)

BACKUP

SENSITIVITY ANALYSES FOR INFORMATIVE CENSORING

- There are a few ways to get some insight

  - Worst-case survival analysis (Allison, 2011) [Allison, P. Survival analysis using SAS: A practical guide, 2nd edition, pp. 283-286].

    - Lower bound: Classify censored cases as having events immediately after censoring

    - Upper bound: Reclassify censored cases as being censored after all events in the sample

    - The actual estimate will be between the upper and lower bound.

    - If neither the upper bound analysis nor the lower bound analysis substantially changes the conclusions, this provides some protection against the impact of informative censoring.

  - Reverse events and censoring

    - Transform data so to reverse events and censoring and perform a time-to-censoring analysis

    - If time to censoring is approximately random, then then censoring times will be approximately evenly distributed and consistent across treatment arms

    - This provides some evidence censoring occurs (completely) at random.

INFORMATIVENESS EXAMPLES

- Common potentially informative reasons for discontinuing follow-up, generally resulting in censoring, include:

  - Patient withdrawn from study due to clinical progression

  - Patient not able (e.g. too ill) to come to clinic or does not want to continue treatment

  - Patient seeking other therapy/entering a new trial

- Likely uninformative reasons for censoring include:

  - Analysis cutoff

  - Administrative closure of study

  - Completion of pre-defined follow-up period.

Most **patient-initiated reasons** for discontinuation have at least **potential for informativeness**

FEASIBILITY OF FOLLOWING PATIENTS INDEPENDENTLY OF TREATMENT IS A QUESTION OF PRACTICALITIES

- May be infeasible or unethical if patients
  - Enter a new drug trial with new experimental therapy after withdrawal
  - Receive frequent/extensive/intrusive/radioactive imaging or similar methods are required (Continuing invasive/risky assessments beyond benefit may not be ethical).
  - Live far away from the clinic
- May be more likely if
  - Treatment safety delays are not long, and/or intermediate safety visits are appropriate
  - Patients receive subsequent and/or supportive therapy at the same clinic
  - The study's required measurements and clinic visits after treatment withdrawal are not much more frequent/intrusive/difficult than under standard-of-care.
  - Patients are informed about and motivated to continue participation
- There are a variety of other practical considerations