Applying the Estimands Framework to Oncology Studies: What Happens When You Cannot Follow Past an Intercurrent Event?

Jonathan Siegel, Bayer

Science For A Better Life

The Pharmaceutical Industry Working Group on Estimands in Oncology

Censoring Mechanisms Subteam

Lifetime Data Science Conference

June 2, 2023

Disclaimer

*The views presented in this presentation are **solely those of the author**, do not necessarily represent the views of the Pharmaceutical Industry Working Group on Estimands in Oncology, Censoring Mechanisms Subteam in every respect, and do not represent the views of the author's employer or any other organization.*

Outline

- Background
- Regulatory and industry context
- Some quick terminology
- The treatment policy estimand
  - Key assumptions
- Alternatives to the treatment policy strategy
  - Strategy selection and feasibility
  - Composite strategy
  - While-on-treatment strategy
  - Hypothetical strategies
  - Principle stratum strategy
- Design as an engineering process
- Summary and conclusions

Background

- Our paper and this presentation assume familiarity with the ICH E9 (R1) Addendum and its principles.
- The Pharmaceutical Industry Working Group on Estimands in Oncology Censoring Mechanisms Subteam has been meeting since 2018 to clarify applicability of the estimands guidance in time-to-event cancer trials.
- In the past, methods in pharmaceutical oncology time-to-event clinical trials have been highly standardized
- The estimands guidance focuses studies on addressing a clinically meaningful, feasible research question
  - It requires justifying assumptions and potentially opens a wider range of tools, particularly tools based on the causal estimands framework (Pearl, 2009).
  - We believe it has analogies to the concept of quality as fitness for purpose (Deming, 1986)

Regulatory and industry context

- Some historical practices have been regulatorily mandated.
  - The 2007 FDA Cancer Endpoint guidance mandated censoring PFS for subsequent therapy and other ICEs.
  - Only superseded in 2018 guidance. Past practice is still optional, still widely done.
- Competitive environment requires rapid development with limited up-front knowledge
  - Accelerated approval permits launching a late-phase TTE study with only a small Phase 1b/2a response study.
  - Designs based on guesswork.
- Trials serve multiple audiences with different goals
  - Regulatory agencies, EU/Canada payer organizations, practicing clinicians, etc., all with different cultural/historical concepts of "treatment effect."

Regulatory/industry context cont.

- Context favors robustness to assumptions
  - New techniques often improve inference by making more assumptions, some traditional criticism is warranted.
- Post-hoc interpretation remains important
  - With limited up-front knowledge and many opportunities for error, flawed trials still need to be interpreted, and may still represent evidence for or against approval.

Some quick terminology

- Implied estimand. The implied estimand (Rufibach, 2019) is the best post-hoc interpretation of the results. It may be different from the estimand intended at design.
- Process termination. Process termination occurs when measurement is stopped prior to reaching the event of interest.
  - Natural process termination occurs when the underlying process is stopped by an intercurrent event, e.g. a terminal event such as death.
  - Artificial process termination occurs when the measurement process is stopped by an event occurring within the trial, e.g. patient withdrawal to enter a new trial.

The Treatment Policy Strategy and its alternatives

The treatment policy strategy is often the one most similar to past conventional practice (although not the same).

The treatment policy strategy

- We start with this strategy because it is closest to the traditional "ITT" approach of the ICH E9 guidance and has generally been preferred by regulators for efficacy trials.
- In the context of treatment switching, the scientific question addresses the effect of the complete regimen including assigned study treatment and all subsequent therapy. Subsequent therapy is part of the treatment element.
- More generally, applied to other intercurrent events, the scientific question is concerned with the outcome from randomization to the event of interest, through and beyond the intercurrent event.
  - The strategy assesses the total effect with the effect of the intercurrent event modeled as part of the treatment effect (Lipkovich, 2022)
  - In a randomized study, it "essentially assesses the effect of being randomized to treatment" (Lipkovich, 2022)

Key assumptions

- Our paper focuses on two critical assumptions underlying the treatment policy strategy:
  - Patients can be followed systematically beyond the applicable intercurrent event
  - The treatment regimen including subsequent therapy predicts future clinical practice in the particular decision context.

Consistent follow-up and its absence

- Implementing this strategy requires an ability to consistently follow patients through and beyond the intercurrent event.
- This is not always feasible.
- A key concern of the paper is what can be done in cases where consistent follow-up is not feasible
  - This is not the ordinary case.
  - But it is an important special case.

Examples where follow-up is infeasible

- Terminal events
  - Patients cannot be followed beyond terminal events, e.g. death.
  - Appropriately accounting for death requires careful consideration in virtually all estimands in an oncology trial.
- Treatment failure
  - Patients may need to enter a new, incompatible trial.
- Open-label trials
  - Within some open-label trials, large fractions of patients randomized to control treatment have left the trial shortly after randomization. (e.g. Larkin 2018; Cortes 2019).
- Functional unblinding
  - Side effects or other signatures have resulted in functionally unblinding some blinded trials, resulting in patient behavior similar to open-label trials.

Process termination and treatment policy strategies

- Terminal events result in natural process termination and tend to defeat implementation of a treatment policy estimand.
  - "In general, the treatment policy strategy cannot be implemented for intercurrent events that are terminal events, since values for the variable after the intercurrent event do not exist." (ICH E9 [R1])
- Our paper focuses on artificial process termination, non-terminal events that, within the context of a study (design and/or conduct), prevent the event of interest from being observed.

Artificial process termination

- Many aspects of the study design and execution can lead to patients being lost to observation at or after intercurrent events
  - Patient management rules, rules for ending treatment and follow-up assessments, unexpected patient behavior, traditional explicit censoring rules, etc.
- A single isolated event can introduce informative censoring but does not alter the overall interpretation of the results.
- Where systematic artificial process termination occurs, however, the effects can be sufficient to change the overall interpretation.
- When this occurs, the implied estimand reflects something other than a treatment policy strategy.
- It’s a more common problem than may be realized.

Predictiveness and its absence

- Behavior of patients in a clinical trial may sometimes be due to the special conditions of the clinical trial environment
- For example, patients may choose subsequent therapy that they would not have chosen if they had taken the study drug in real-world clinical practice
- Examples:
    - Subsequent therapy is itself experimental and unavailable as SOC.
    - Patients assigned to control withdraw study treatment and receive subsequent therapy immediately or early.

Example: Immediate treatment switching in open label trials

Open-label studies have the risk that patients stop randomized treatment after randomization in the control arm and seek the opportunity to receive an investigational therapy in another clinical trial, possibly even from the same class as the investigational drug in the previous trial.

Examples:

- Checkmate-37 (Larkin, 2018) comparing Nivolumab vs chemotherapy where 20% of the patients from the control arm withdrew consent immediately after they learned that they were randomized into the control arm (vs. 1.5% on investigational arm)
- Quantum-R trial (Cortes, 2019): 23% in placebo withdrew immediately vs 1.6% on investigational arm

Larkin J., Minor D., D'Angelo S., Neyns B., Smylie M., Miller W.H. Jr., Gutzmer R., Linette G., Chmielowski B., Lao C.D., Lorigan P., Grossmann K., Hassel J.C., Sznol M., Daud A., Sosman J., Knushalani N., Schadendorf D., Hoeller C., Walker D., Kong G., Horak C., Weber J., Overall Survival in Patients With Advanced Melanoma Who Received Nivolumab Versus Investigator's Choice Chemotherapy in CheckMate 037: A Randomized, Controlled, Open-Label Phase III Trial. J Clin Oncol. 2018; 36(4):383–390. doi:10.1200/JCO.2016.71.8023

Alternatives to Treatment Policy

We discuss alternatives to the treatment policy strategy that study teams should consider when the assumptions underlying the treatment policy strategy are particularly likely to break down.

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, highly related to, or worse than 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
- A composite strategy has some important advantages
  - Does not require follow-up beyond the intercurrent event
  - Clear statistical validity without making questionable assumptions
  - Often (not always) clinically meaningful and relevant

COMPOSITE STRATEGY EXAMPLE: PROGRESSION

- Progression-free survival, a composition of time to progression and overall survival, is a well-known oncology example
- A composite strategy might sometimes be considered for clinical progression.
  - In some cases, patients may be particularly likely to end tumor assessments at clinical progression without waiting for formally documented progression.
  - Clinical progression has its own reliability, interpretation, and bias issues.
    - It has generally not been recommended for regulatory approval for these reasons
  - In particular cases where patients systematically do not continue tumor assessments, bias due to informative censoring resulting from ignoring clinical progression may outweigh bias due to unreliability in assessing it.

WHILE ON TREATMENT STRATEGY

- The "while on treatment" strategy poses a research question that is only interested in the treatment effect until the intercurrent event occurs.
- The strategy can be applied to any intercurrent event (not just treatment withdrawal)., where the strategy could be described as while-prior to occluding event.
  - A particularly important example is death ("while alive").
- 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 treatment failure for a palliative treatment TTE analysis.

While-on-treatment strategy implementation: Independent causes

- Unkel (2019) has characterized standard TTE estimation (Kaplan-Meier, Cox models, etc.) as implementing an independent-causes while-on-treatment strategy.
- Rufibach (2019) has characterized it as a hypothetical strategy.
  - We think this the better interpretation..
- We do not generally recommend this approach, regardless of how characterized or interpreted.
  - Will discuss further when discussing hypothetical strategies.

Do CIF and Fine-Gray methods implement a while-on-treatment strategy?

- Not everybody agrees. Perhaps competing risks implement a separate strategy.
- We think that of the guidance strategies, while-on-treatment is the best fit (although somewhat imperfect), at least for CIF.
- It is not a treatment policy strategy. It can be readily used for terminal events like death. Treatment policy strategies can't be.
- It is not concerned with incidence after the ICE.
- It has the essential weakness of a general while-on-treatment strategy identified by Han and Zhou (2023).
  - In Han and Zhou's palliative example, "If the pain for patients with chronic diseases increases with time, a poisonous drug that can kill people in a relatively short time could produce better results than a placebo, which is misleading."
  - CIF (and Fine-Gray) have just this problem.

CIF and Fine-Gray model issues

- In the Fine-Gray model, hazards are dependent, which generally precludes a causal interpretation.
- The CIF approach depends on the incidence, not the hazards.
  - It preserves causal interpretability, but is largely a descriptive rather than an inferential method.
- The Han and Zhou weakness is a significant limitation.
  - To address it, an estimand based on the underlying intercurrent event could be evaluated first.
    - Example In a trial with overall survival as primary objective, death could be used a competing risk for secondary objectives only after survival superiority or non-inferiority has been established.
  - Otherwise, a palliative context is critical
    - For a strictly palliative drug, we don’t care about survival
    - But we generally do care.

Hypothetical strategies

- A hypothetical strategy addresses a counterfactual scientific question, generally what would have happened if an intercurrent event had not occurred.
  - As we've discussed, when assessments systematically end following an intercurrent event, simple censoring induces systematic artificial process terminations, altering the implied estimand.
  - We believe this induces an implied hypothetical strategy
    - It reflects what efficacy would be if the intercurrent event had not occurred, and also if subsequent hazards remained the same as in patients in whom it did not occur.
  - Making this strategy explicit at study design would represent an improvement over leaving it implicit.
  - But we rarely recommend doing this in oncology, as the non-informativity assumption is often questionable.

Hypothetical strategies and causal inference

- Causal-inference hypothetical strategies predict the behavior of patients experiencing the intercurrent event using e.g. propensity scores based on the behavior of “similar” patients who did not.
- The implementing methods make strong assumptions
  - In particular, they assume that all systematic effects have been modeled.
  - This can be a dubious assumption in oncology
  - We often can’t even predict which classes of patients will benefit from treatment and which won’t.
    - If our model can’t even account for that, how can it incorporate all systematic effects?
- We identify a narrow set of cases where a causal-inference hypothetical strategy might be considered, despite these limitations.
- See Manitz et al. (2022), the Estimands WG treatment switching paper, for more details.

When the trial induces non-predictive behavior

- As mentioned at the beginning of this presentation, in some circumstances a trial can induce behavior that would not be observed by patients in a clinic outside the trial context.
- In these cases, randomization and blinding is inducing patient behavior that would not occur if randomization and blinding had not happened.
  - The trial introduces a non-predictive environment
- We argue a treatment policy strategy is answering the wrong question.

A hypothetical strategy answers the right question

- A causal-inference hypothetical strategy asks what would have happened if the not-predictive-of-the-clinic trial behavior hadn't happened.
  - We fully agree that in general it is not clinically meaningful to answer hypothetical questions.
  - However, for certain extreme deviations from treatment policy assumptions, the strategy could be more predictive of what would have happened if patients had been in a clinic instead of in the trial.
- A causal hypothetical strategy in this situation may not provide a fully reliable answer.
  - But, in appropriate special circumstances, it answers the right question.
  - We believe addressing the right question carries weight in an estimands environment.
- In a situation where randomized trials will inevitably produce non-predictive behavior and a trial that does not is simply not feasible, this may be the only way to approve entire classes of useful drugs.

PRINCIPAL STRATUM STRATEGY

- 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 will not occur
- Classification is modeled based on characteristics known at baseline
- Issues
  - Modeling for principle stratum requires strong assumptions.
  - For a number of important intercurrent events, identifying the treatment effect in patients in whom the intercurrent event does not occur may not address a clinically meaningful question.
- Has not had a lot of use in oncology clinical trials
- See Bornkamp et al (2021), a paper by the WG Causal Estimands Task Force, for further information on the potential role of this strategy in drug development.

Engineering Approach

In particular clinical trial conditions, it may sometimes be necessary to trade off or compromise between optimal scientific appropriateness and feasibility

We recommend an engineering approach to these issues

Addressing conflicting needs

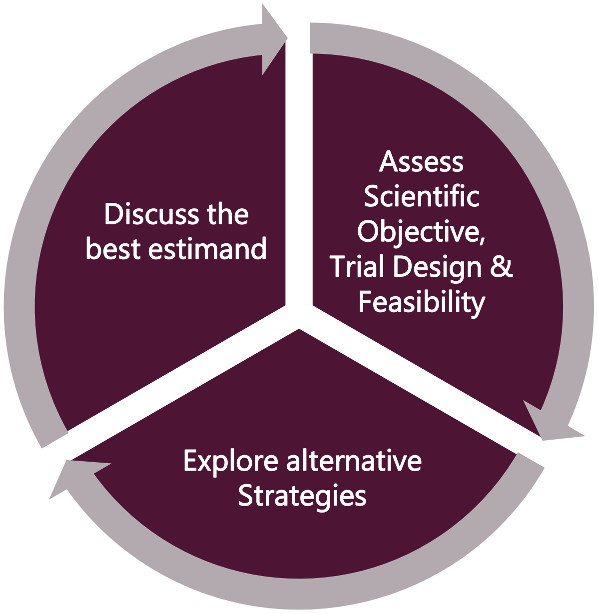
- Oncology clinical trials often have to address conflicts among different requirements
  - Dosing, patient management, and ethical-care needs often drive clinic visit schedules and clinical assessment termination requirements
  - Higher-priority estimands may drive what little flexibility is left.
  - Secondary estimands may have to accept a visit schedule ending assessments based on the needs of patient management and the primary estimand.
- Where assessments can continue by alternative means, through e.g. electronic diaries, this should be done.
  - But this is not always possible.
- Teams should identify these cases and specify estimands that can be feasibly addressed by the visit schedule.

Study design as a feedback loop

- In practice, feasibility and ethical considerations, the needs of other estimands, and other factors will often require revisiting what can be feasibly addressed in the study context.
- As a result, the design process will often involve a feedback loop analogous to Deming's Plan-Do-Study-Act cycle (Deming, 1986).
- Study teams will need to collaborate with a clear understanding of goals and clinical and operational conditions.
- It is important to start with the ideally desired research question, not just what appears feasibly implementable.
  - It is always possible that design changes, technology, and other improvements may overcome constraints and enable the original research question to be addressed.

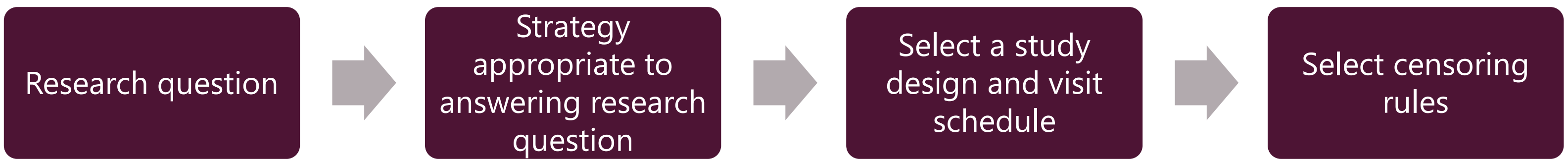
STRATEGY SELECTION AND FEASIBILITY

- 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.

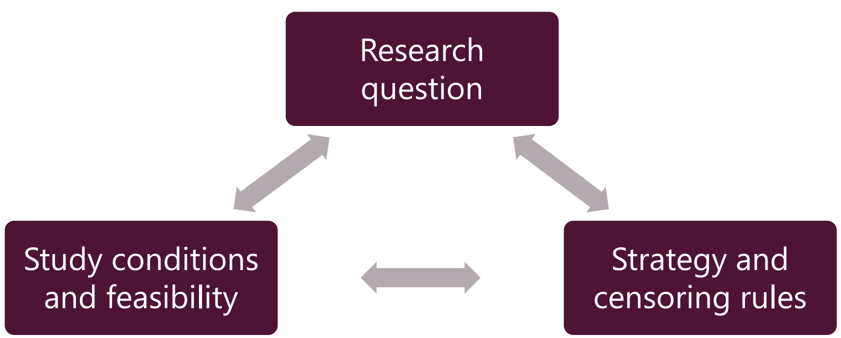


STUDY DESIGN MAY BECOME A LESS LINEAR PROCESS

- Classical causal estimands theory describes study design as a linear process:



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



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

Summary and conclusions

- The estimands framework requires rethinking long-standing habits in time-to-event oncology trials
  - Careful consideration of research goals and strategies
  - Understanding context, constraints, and alternatives.
- Systematic process termination is inconsistent with a treatment policy strategy
  - An alternative strategy should be considered in consultation with the study design team.
  - It happens more often than might be realized.
- The complex clinical context of oncology, with non-proportional hazards, correlations among outcomes, discrete assessments, ethical and patient management requirements, etc., makes designing and interpreting time-to-event estimands particularly challenging.
- Cross-discipline cooperation and an engineering approach is recommended.

REFERENCES

- Anderson, P. Censored Data. *Encyclopedia of Biostatistics, 2nd Edition.* Wiley (2005). https://doi.org/10.1002/0470011815.b2a11008
- Carroll KJ (2007): Analysis of progression-free survival in oncology trials: Some common statistical issues. Stats Med 6: 99-113
- Cortes, J et al. Quizartinib versus salvage chemotherapy in relapsed or refractory FLT3-ITD acute myeloid leukaemia (QuANTUM-R): a multicentre, randomised, controlled, open-label, phase 3 trial. Lancet 20:984-97 (2019)
- Deming, WE. *Out of the Crisis*. MIT Press (1982)
- Fine J, Gray R (1999); A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc. 94:496-509
- 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)
- Han, S and Zhou, X. Defining estimands in clinical trials: A unified procedure. Stat Med. 42:1869-87 (2023)
- Hernan, M & Robins, J. *Causal Inference: What If.* CRC Press (2021). See Fine Point 17.1, p. 210.
- Kurland, B. et al. Longitudinal data with follow-up truncated by death: Match the analysis method to research aims. *Stat Sci.* 24:211–222 (2009)

REFERENCES (CONT.)

- Larkin J. et al .Overall survival in patients with advanced melanoma who received nivolumab versus investigator's choice chemotherapy in CheckMate 037: A randomized, controlled, open-Label Phase III trial. J *Clin Oncol*. 36(4):383-390 (2018)
- Lipkovich, I. et al. Causal inference and estimands in clinical trials. , *Stat Biopharm Res.* 12:1, 54-67 (2020)
- Pearl, J. *Causality: Models, Reasoning, and Inference 2nd Edition*. Cambridge University Press (2009)
- Rubin, D. Causal inference through potential outcomes and principal stratification: Application to studies with 'censoring' due to death. *Stat Sci* 21:299-309 (2006)
- Rufibach, K. Treatment effect quantification for time-to-event endpoints–Estimands, analysis strategies, and beyond. Treatment effect quantification for time-to-event endpoints–Estimands, analysis strategies, and beyond. Pharmaceutical Statistics 18:145-165 (2019)
- Schuster, SJ et al. Tisagenlecleucel in adult relapsed or refractory diffuse large B-cell lymphoma *N Engl J Med* 380:45-56 (2019)
- Unkel, S. et al. (2019). On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. *Pharm Stat* 18:166–183.
- US H&HS FDA. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics: Guidance for Industry (2007).
- US H&HS FDA. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics: Guidance for Industry (2019).
- US H&HS FDA. E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials: Guidance for Industry (2021)
- Yang, F & Ding, P. Using survival information in truncation by death problems without the monotonicity assumption. *Biometrics* 74:1232-9 (2018)

Thank you!