Estimands in Oncology: New Directions

Qing Xu, FDA

Jonathan Siegel, Bayer

Libby Floden, Clinical Outcomes Solutions

Janet Wittes, Wittes LLC

Hongtao Zhang, Merck

Chair: Yeh-Fong Chen, FDA

Panel Discussion, BIOP-RISW 2023 Conference, September 29, 2023

On behalf of the ASA Biopharmaceutical Section Estimands in Oncology Scientific Working Group

Disclaimer – General and FDA

The comments expressed herein are the authors' own and should not be interpreted in any way as representing their respective employers or FDA's views or policies.

U.S. FOOD & DRUG ADMINISTRATION

Question #1: What are appropriate Estimand strategies to handle potential informative censoring in time-to-event endpoints?

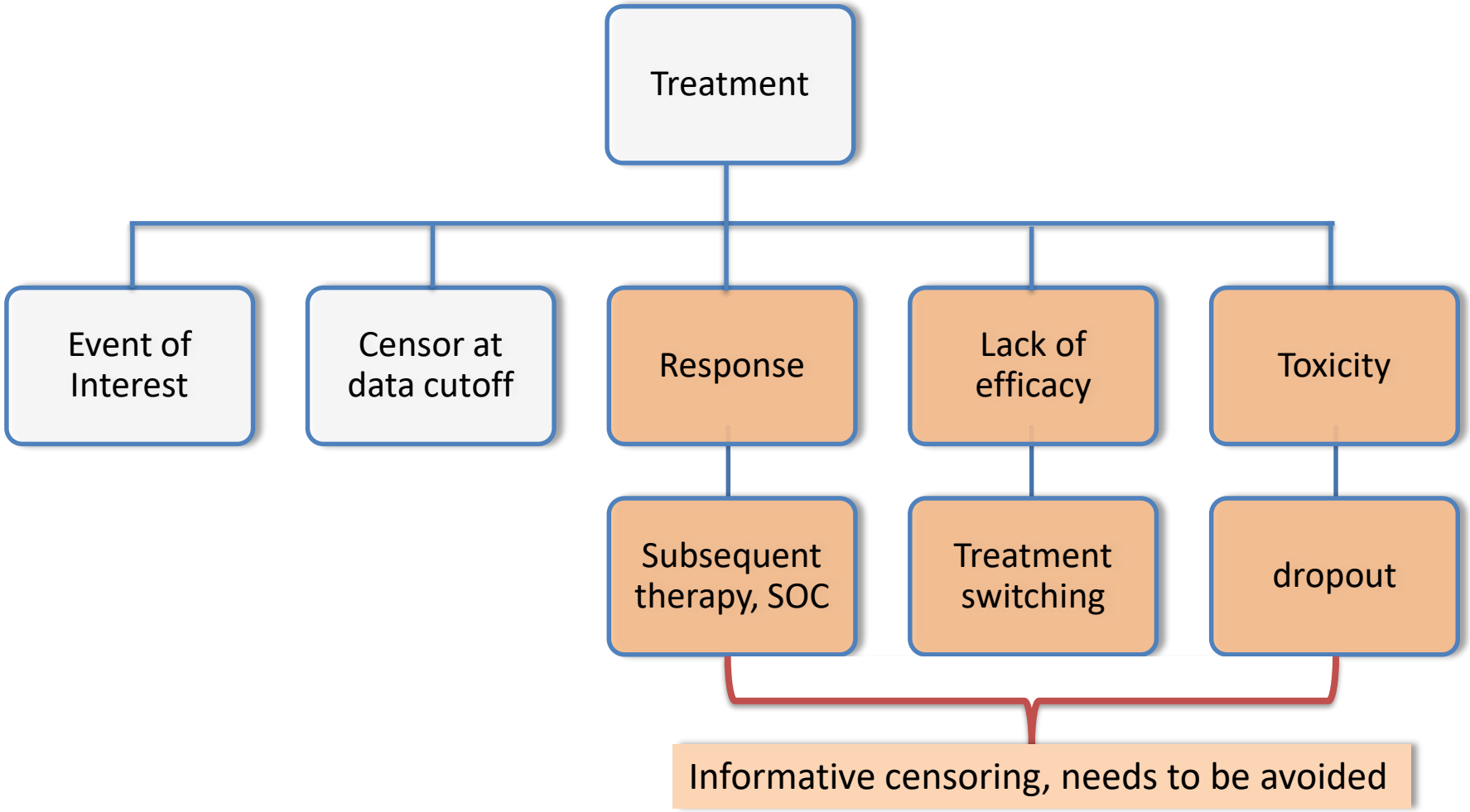
Panelist: Qing Xu, Ph.D

Key Components of Survival Analysis

- **Proportional hazard assumption**

- **Non-informative censoring**
  - Subjects who drop out of the study or receive subsequent therapy should do so for reasons unrelated to the treatment
  - Censoring distribution is unrelated to the event or other related variables
  - Ideally, the subjects who are censored at time t should be representative of all the subjects who remained at risk at that time with respect to their survival experience
- **Informative censoring may lead to a biased estimate**

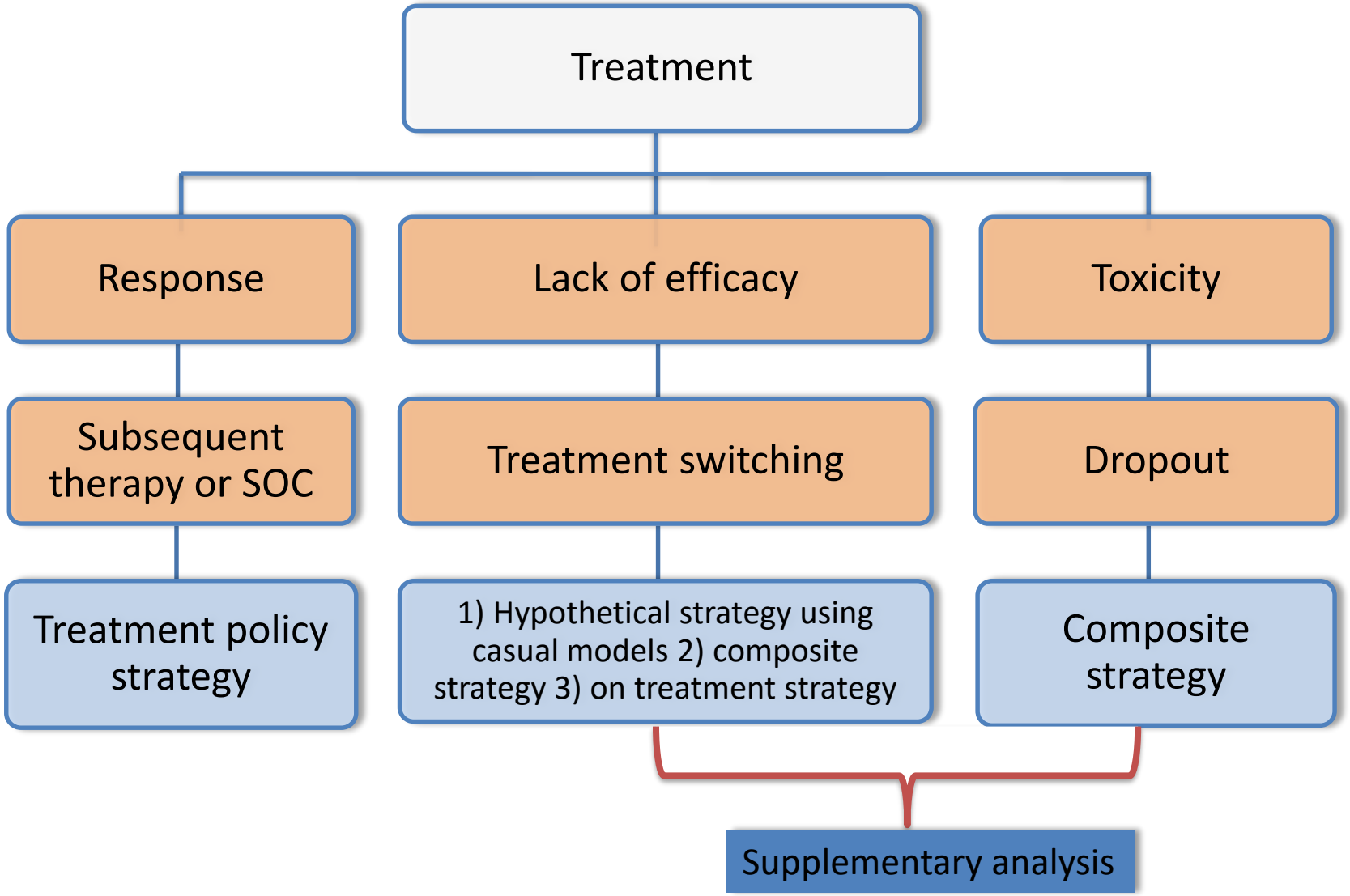
Survival Analysis Treatment Scheme



Estimand Strategies for Handling Intercurrent Events

- **Treatment policy strategy**
  - The occurrence of the intercurrent event is considered irrelevant in defining the treatment effect of interest
- **Hypothetical strategies**
  - A scenario is envisaged in which the intercurrent event would not occur
- **Composite variable strategies**
  - Intercurrent event is considered in itself to be informative about the patient's outcome and is therefore incorporated into the definition of the variable.
- **While on treatment strategies**
  - Response to treatment prior to the occurrence of the intercurrent event is of interest.
- **Principal stratum strategies**
  - Restrict the population of interest to the stratum of patients in which an intercurrent event would not have happened

Estimand in Intercurrent Events



PFS Analysis: Polatuzumab ODAC

POLARIX is a multicenter, randomized, double-blind, placebo-controlled trial comparing the efficacy and safety of pola+R-CHP to R-CHOP in 879 adult patients with untreated large B-cell lymphoma

PFS Analyses by Censoring Rules	Difference in 2-year PFS	HR (95% CI)	p-value
Original Data			
NALT: Not Censor	6.5%	0.73 (0.57, 0.95)	0.0177
Sensitivity Analyses on Original Data (nominal p-values)			
NALT: Censor	4.9%	0.77 (0.59, 1.01)	0.0541

NALT: New anti-lymphoma therapy

Source: Polatuzumab ODAC Presentation

Summary

- **There is no satisfactory way to correct for informative censoring**
  - Capture the reasons why patients started new therapies and how the new therapies will affect the outcomes
  - Capture the reasons of dropout

- **Different methods with different Estimand strategies should be conducted to ensure robustness. No single Estimand that fits all needs**

Question #2

September 24, 2021

It is not always possible to follow patients until the event of interest. How can this be addressed during study planning as well as during analysis and reporting?

Panelist: Jonathan Siegel, Bayer

The treatment policy strategy

- The treatment policy strategy is
  - Closest to the traditional "ITT" approach
  - Often preferred by regulators
- Scientific question concerns outcome from (e.g.) randomization to the event of interest, through and beyond the intercurrent event.
  - Attributes effect of intercurrent event to treatment effect
  - Requires patients to be followed systematically through and beyond the applicable intercurrent event
  - Assumes clinical trial outcomes predict future clinical practice
- Systematic inability to follow patients generally defeats a treatment policy strategy
- As many oncology assessments depend heavily on clinic visits, consistent follow-up is not always possible

Examples of large-scale/highly informative loss to follow-up

- Terminal events, e.g. death.
- Treatment failure: Patients enter a new, incompatible trial and cannot continue in the existing one.
- Open-label trials: Large fractions of patients randomized to control leave the trial shortly after randomization.
  - Checkmate-37 (Larkin, 2018, Nivolumab vs chemotherapy) 20% of control-arm patients withdrew consent immediately after being randomized into the control arm, vs. 1.5% on investigational arm[1]
  - Quantum-R trial (Cortes, 2019): 23% in placebo withdrew immediately vs 1.6% on investigational arm[2]
- Functional unblinding: side effects can reveal treatment
- Withdrawal criteria: When treatment withdrawal or progression end clinic visits, they stop follow-up for all other estimands requiring it.

[1]Larkin J. et al. *J Clin Oncol.* 36(4):383-390 (2018) [2]Cortes, J et al. *Lancet* 20:984-97 (2019)

What can be done

- Where possible, visit schedule, protocol instructions, informed consent, etc. should be structured to ask patients to continue follow-up past treatment withdrawal, subsequent therapy, etc.
- It may be possible to follow patients in a different way
  - Phone calls
  - On-line
  - Personal devices
  - House visits
- Where not possible, alternative strategies should be considered.
- It is often necessary for high priority estimands (e.g. progression, toxicity) to control the visit schedule and end clinic visits needed for other, subordinate estimands
  - When this is unavoidable, informative loss to follow-up should be explicitly addressed with an appropriate strategy.
  - The trial context often requires compromises from the ideally desired estimand.

Strategies not requiring further follow-up

- Composite strategy
  - Makes the ICE a component of the event of interest.
    - Pervasive example: PFS (composition of progression and death)
  - Requires few assumptions. Generally least problematic for inference
  - Not always clinically meaningful

While on Treatment Strategy

- Can be applied to any ICE, not just treatment (e.g. “while alive”)
- Only concerned with what happens up to the time of the ICE
- May be clinically inappropriate.
  - Especially for ICEs like death or progression, higher incidence of ICE may reduce apparent incidence of event of interest.
  - Events like treatment withdrawal (or progression) may be highly informative of subsequent events,
    - Ignoring may yield misleading results. (Yang et al., 2018)[1]
- Evaluating the ICE as its own endpoint first (e.g. testing for survival earlier in a FWERC hierarchy before using a while-on-treatment strategy for death) provides some protection against this concern (Siegel, 2023) [2]

[1] Yang, F., Wittes, J., and Pitt, B. *Clin Trials* 16:63-70 (2019) [2]Siegel et al. Submitted *to Pharm Statist* (2023) https://arxiv.org/abs/2203.01781

Hypothetical Strategy

- Scientific question is what would have happened if the ICE had not occurred.
  - Implementations based on causal inference generally require strong and highly questionable assumptions
    - E.g. assumption of no unmeasured confounders
  - Not generally accepted by FDA.
- We (Siegel, 2023[1]; Manitz, 2022[2]) argue for regulatory consideration of a hypothetical strategy under narrow circumstances
- Let's go back to our example where a large fraction of patients switches treatments.
  - Why does this happen?
  - The trial context is inducing patient behavior that would not be observed by patients in a clinic outside the trial context.
    - Patients get assigned to treatment they don't really want.

[1]Siegel et al. Submitted *to Pharm Statist* (2023) https://arxiv.org/abs/2203.01781 [2]Manitz J et al. *Pharm Statist* 2022; 21:150-162.

When a trial does not predict the clinic

- The trial does not predict the clinic
  - In the clinic, patients would rarely immediately switch from assigned treatment
- When trial behavior does not predict clinic behavior, we suggest a treatment policy strategy is answering the wrong question
- A hypothetical strategy asks what would have happen if this specific non-predictive behavior had not occurred.
  - It asks the right question. It asks what would have happened in the clinic, the context of real scientific interest.

Clinically meaningful questions vs. reliable answers

- The ICH E9 (R1) guidance requires asking clinically meaningful questions and answering them in a scientifically reliable way.
  - This example illustrates the tension between the two
- We think it is sometimes necessary to compromise between clinical meaningfulness and scientific reliability.
- Asking the right question carries weight, even if the answer cannot be completely reliable.
  - When we lose our keys in a dark alley, looking under the more-reliable streetlamp is not always the wisest course.[1,2]
- We suggest a hypothetical strategy should be considered where a clinical trial induces systematic patient behavior that will not predict the clinic if we simply ignore it.
- The clinically meaningful question, what we really want to know, is what would have happened if that behavior had not occurred.
- Where pervasive non-predictive behavior cannot be avoided, it may be the only way to evaluate entire classes of drugs.

[1]Freedman, D., The Streetlight Effect. *Discover* (August 1, 2010). [2]Goldstein, Z. *The Silly World of Chelm*. Two Lights Publishing (2016)

Final note – Compromises, cooperation, communication

- This brief introduction illustrates the importance of devising good compromises between scientific and clinical/contextual considerations
  - Need to address tension between the two forthrightly, with eyes wide open, understanding both.
  - As good compromises require expert knowledge in multiple fields, cooperation and clear communication among experts is essential.

Thank you!

Are there unique challenges to implementing the estimand framework for patient-reported endpoints?

What are they, and what are potential strategies for handling these challenges?

Libby Floden, PhD MPH
Senior Director, Quantitative Science
Clinical Outcomes Solutions

Discussion points derived in the manuscript, currently under review:
Rachael Lawrance, Konstantina Skaltsa, Antoine Regnault, Lysbeth Floden, “Reflections on estimands for patient-reported outcomes in cancer clinical trials”

Estimand Attribute: Variable of interest

- Example of a (poorly defined) PRO variable: “Evaluate health-related quality of life”
- Concept of interest: PRO endpoints are meant to capture concepts that are not necessarily obvious to clinicians and that therefore need to be explicitly defined in the first place
  - Symptoms (tolerability)
  - Physical functioning
- Refinement process
  - Relevant to trial population
  - Appropriate for study design
  - Specific to the concept of interest
  - Measurable via a PRO instrument

Estimand Attribute: Variable of interest

Comparing two groups (randomized arms, cohorts, pre-post, …)

- We found a **difference in quality of life**
- We found a **difference in self-reported physical functioning**
- We found a **difference in self-reported physical functioning at month 6**
- We found a **difference in mean self-reported physical functioning change from baseline at month 6**
- We found a **difference in mean self-reported physical functioning change from baseline at month 6 for patients still on-treatment.**

Adapted from Floden, Roydhouse, Lawrance, and Coens. “Patient-reported outcomes in clinical trials - what questions are we asking and how do we understand the results; how the estimand framework with sensitivity analyses can help” Workshop presented at ISOQOL Annual Conference, October 2022

Estimand Framework: Missing data vs Intercurrent Event

- ICEs: Post-randomisation events that affect the measurement or interpretation of a variable
- Missing data: an outcome value that is meaningful for analysis was not collected
  - Patient A assigned to investigational arm - withdraws due to toxicity
    - Post-withdrawal values may be meaningful → missing
  - Patient B assigned to control arm - dies
    - Post-death values not meaningful → not missing
    - But death is an important post-randomisation consideration
- Rombach et al review: ~40% included studies had information on patients with PROMs at main follow-up point[1]

1. Rombach, I., Rivero-Arias, O., Gray, A.M., Jenkinson, C., Burke, O. (2016). The current practice of handling and reporting missing outcome data in eight widely used PROMs in RCT publications: a review of the current literature. *Quality of Life Research 25:* 1613-1623

ICEs

- PRO endpoints in (oncology) clinical trials are most often one the three following types:
  - Magnitude of change,
  - Time-to-event, and
  - Proportion of responders[1]
- ICEs can include the primary trial outcome!
  - Need to consider carefully in the evaluation of PRO endpoints
  - Often represent competing events

1. Coens, C., M. Pe, A. C. Dueck, J. Sloan, E. Basch, M. Calvert, A. Campbell, C. Cleeland, K. Cocks and L. Collette (2020). "International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium." *The Lancet Oncology* **21**(2): e83-e96.

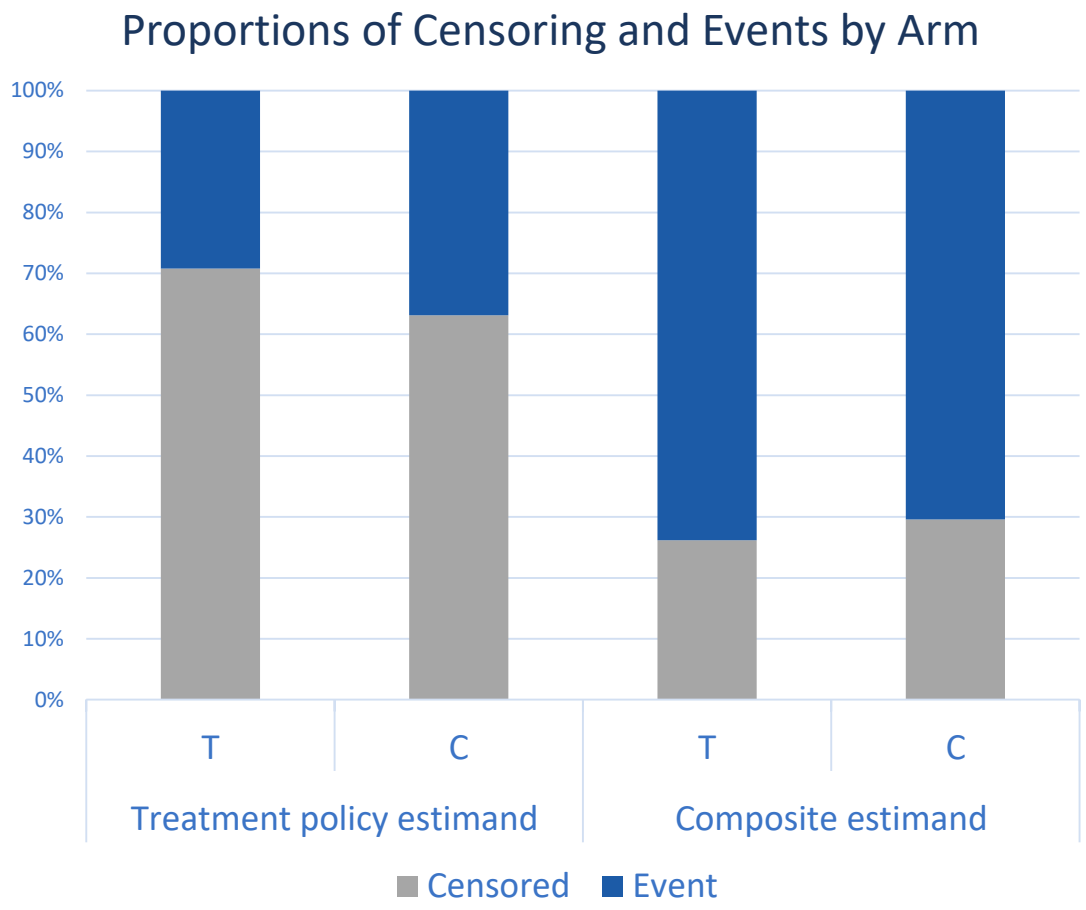
Strategies: Treatment policy challenges

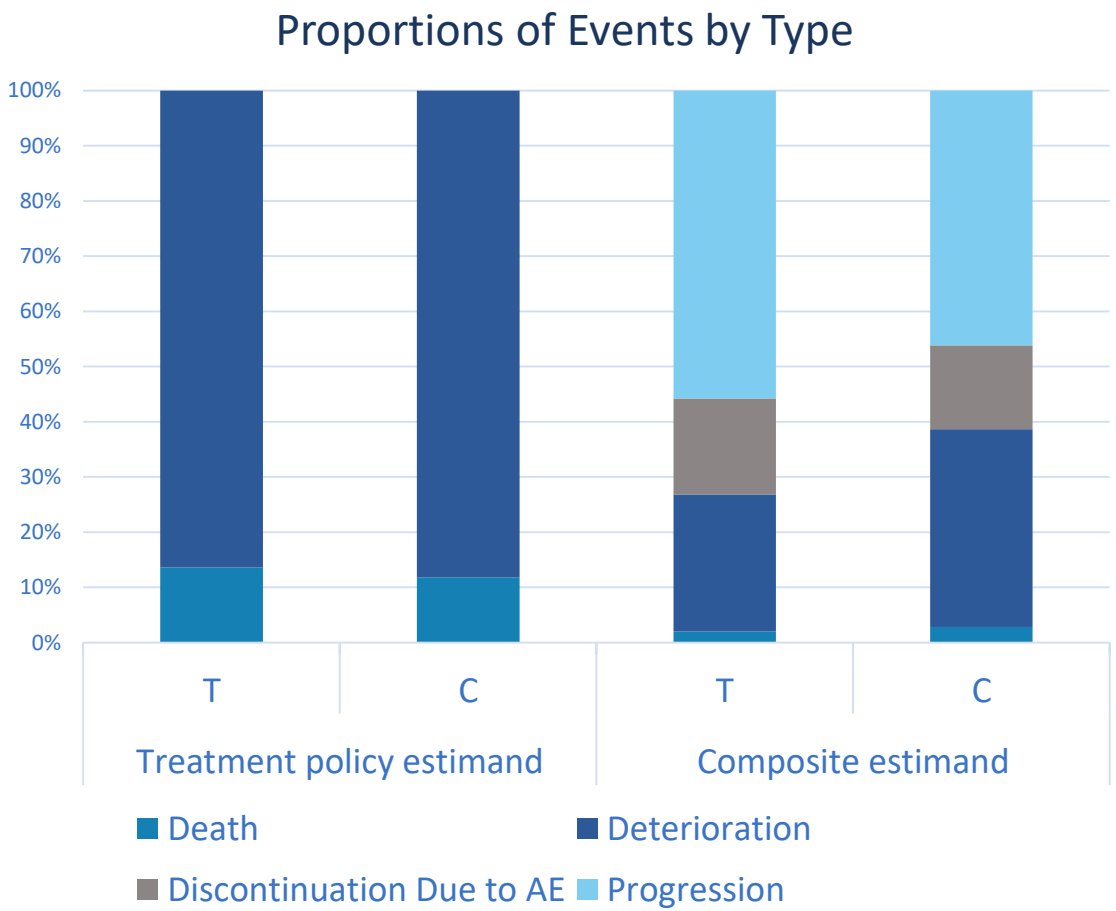
- Embraced as the closest-to-the-ITT-principal approach
- Despite value of PRO data across the patient's journey, post-discontinuation PRO collection can be:
  - burdensome for the patient, and
  - challenging (and costly) operationally.
- Protocol schedules often stop data collection after treatment discontinues
  - Implication: estimation of treatment effect is reliant on strong statistical assumptions

Strategies: Composite challenges

- Done by dichotomizing a PRO score and combining with other ICEs, e.g., disease progression
- May be an attractive strategy when other endpoints are also ICEs
- But can also make interpretation difficult
  - PRO value and ICE contribute equally to the variable
  - Can consider rank-based methods, e.g.,rank-based ANCOVA, win ratio

Comparing PRO Estimands for TTE





HRs = 0.58 and 0.66 for Treatment and Composite estimands, respectively

Joint work with Mike DeRosa, Jen Beaumont, Jessica Roydhouse, and Stacie Hudgens

Question 4:
Will the estimands framework make any changes in the way we evaluate safety?

Janet Turk Wittes, PhD
ASA-Biopharm Workshop
29-September-2023

Summary 3 point answer

1. Safety is very hard – much harder than efficacy
2. Yes, the estimand framework will lead to changes in approach
3. But many of the changes will be non-rigorous
    - Are we using a can opener to open a peanut butter jar?

1. Why is safety (in onology trials) difficult?

- Many potential harms are not prespecified
- Many oncology trials (even Phase 3) are open-label
  - Arms may have different schedules of treatment and visits
  - Even worse, the length of treatment may differ
  - Control often crosses over to active at progression
- There are all sorts of intercurrent events
  - Deaths
  - Stopping study drug
  - Etc.
- Participants on many drugs; hard to point to experimental drug

2. Yes, estimand framework will lead to changes in approach

- E.g., typical analysis

Table 11: Clinically Important Treatment-Emergent Adverse Reactions (Regardless of Relationship) in Patients with SCCHN Receiving Induction Chemotherapy with Docetaxel in Combination with Cisplatin and Fluorouracil Followed by Radiotherapy (TAX323) or Chemoradiotherapy (TAX324)

	TAX323 (n=355)				TAX324 (n=494)			
	Docetaxel arm (n=174)		Comparator arm (n=181)		Docetaxel arm (n=251)		Comparator arm (n=243)	
Adverse Reaction (by Body System)	Any %	Grade 3/4 %	Any %	Grade 3/4 %	Any %	Grade 3/4 %	Any %	Grade 3/4 %
Neutropenia	93	76	87	53	95	84	84	56
Anemia	89	9	88	14	90	12	86	10
Thrombocytopenia	24	5	47	18	28	4	31	11
Infection	27	9	26	8	23	6	28	5
Febrile neutropenia*	5	N/A	2	N/A	12	N/A	7	N/A
Neutropenic infection	14	N/A	8	N/A	12	N/A	8	N/A
Cancer pain	21	5	16	3	17	9	20	11
Lethargy	41	3	38	3	61	5	56	10

Changing denominator to time on drug doesn't fix everything!

It reminds us: when we create a "statistic" we need to understand what we are estimating

3. But many of the changes will be non-rigorous

- If we rely on ICH E9 R(1), we are permitted to do analyses that violate randomization
- On-treatment analyses – don't get me started
  - Can over- or under-estimate rate
    - E.g., Yang, Wittes, Pitt (2019). Clinical Trials, 16: 63-70
- Hypothetical strategy – can give us a "what if" or "even if" but don't take the estimate seriously
- Principal stratum – post-randomization subgroups

Summary

- Using estimand framework for safety makes us think (good)
- Can get seriously invalid analyses (bad)
- Fundamental problem: studies not designed for safety

Question 4:
Will the estimands framework make any changes in the way we evaluate safety?

Janet Turk Wittes, PhD
ASA-Biopharm Workshop
29-September-2023

Summary 3 point answer

1. Safety is very hard – much harder than efficacy
2. Yes, the estimand framework will lead to changes in approach
3. But many of the changes will be non-rigorous
    - Are we using a can opener to open a peanut butter jar?

1. Why is safety (in onology trials) difficult?

- Many potential harms are not prespecified
- Many oncology trials (even Phase 3) are open-label
  - Arms may have different schedules of treatment and visits
  - Even worse, the length of treatment may differ
  - Control often crosses over to active at progression
- There are all sorts of intercurrent events
  - Deaths
  - Stopping study drug
  - Etc.
- Participants on many drugs; hard to point to experimental drug

2. Yes, estimand framework will lead to changes in approach

- E.g., typical analysis

Table 11: Clinically Important Treatment-Emergent Adverse Reactions (Regardless of Relationship) in Patients with SCCHN Receiving Induction Chemotherapy with Docetaxel in Combination with Cisplatin and Fluorouracil Followed by Radiotherapy (TAX323) or Chemoradiotherapy (TAX324)

	TAX323 (n=355)				TAX324 (n=494)			
	Docetaxel arm (n=174)		Comparator arm (n=181)		Docetaxel arm (n=251)		Comparator arm (n=243)	
Adverse Reaction (by Body System)	Any %	Grade 3/4 %	Any %	Grade 3/4 %	Any %	Grade 3/4 %	Any %	Grade 3/4 %
Neutropenia	93	76	87	53	95	84	84	56
Anemia	89	9	88	14	90	12	86	10
Thrombocytopenia	24	5	47	18	28	4	31	11
Infection	27	9	26	8	23	6	28	5
Febrile neutropenia*	5	N/A	2	N/A	12	N/A	7	N/A
Neutropenic infection	14	N/A	8	N/A	12	N/A	8	N/A
Cancer pain	21	5	16	3	17	9	20	11
Lethargy	41	3	38	3	61	5	56	10

Changing denominator to time on drug doesn't fix everything!

It reminds us: when we create a "statistic" we need to understand what we are estimating

3. But many of the changes will be non-rigorous

- If we rely on ICH E9 R(1), we are permitted to do analyses that violate randomization
- On-treatment analyses – don't get me started
  - Can over- or under-estimate rate
    - E.g., Yang, Wittes, Pitt (2019). Clinical Trials, 16: 63-70
- Hypothetical strategy – can give us a "what if" or "even if" but don't take the estimate seriously
- Principal stratum – post-randomization subgroups

Summary

- Using estimand framework for safety makes us think (good)
- Can get seriously invalid analyses (bad)
- Fundamental problem: studies not designed for safety

Will the Estimands Framework Have Any Impact on Early-Phase Studies?

Panelist: Hongtao Zhang,

on behalf of Early Development Estimand Nexus (EDEN) working group

Scope of Discussion

- Focus on oncology drug development
- Early phase studies refer to:

1. Phase 1a dose-escalation studies
   - Mercier et al. (2023+) Estimands in oncology early clinical development: Assessing the impact of intercurrent events on the dose-toxicity relationship *(under review)*
2. Phase 1b or 2 single arm studies
   - Englert et al. (2023) Defining estimands for efficacy assessment in single arm phase 1b or phase 2 clinical trials in oncology early development, *Pharm Stat*
   - Applicable to studies with multiple single arm cohorts

Why Estimands in DE Studies?

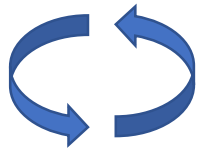
- By having a precise definition of the initial target of estimation, sponsors of ongoing DE clinical trials can better anticipate potential sources of bias due to ICEs.
- Gather feedback to guide statistical design.
- General practice is to replace the patients with the ICE, which is prone to selection bias. This practice could be challenged in view of estimands framework.

Iterative and Adaptive Nature of DE Studies

Objective

Target of decision (not directly estimated)

- Determine the MTD
- Select a dose for next cohort



Estimated quantity

Target of estimation (subject to estimand's definition)

- Pr(DLT)
- Pr(Target toxicity)

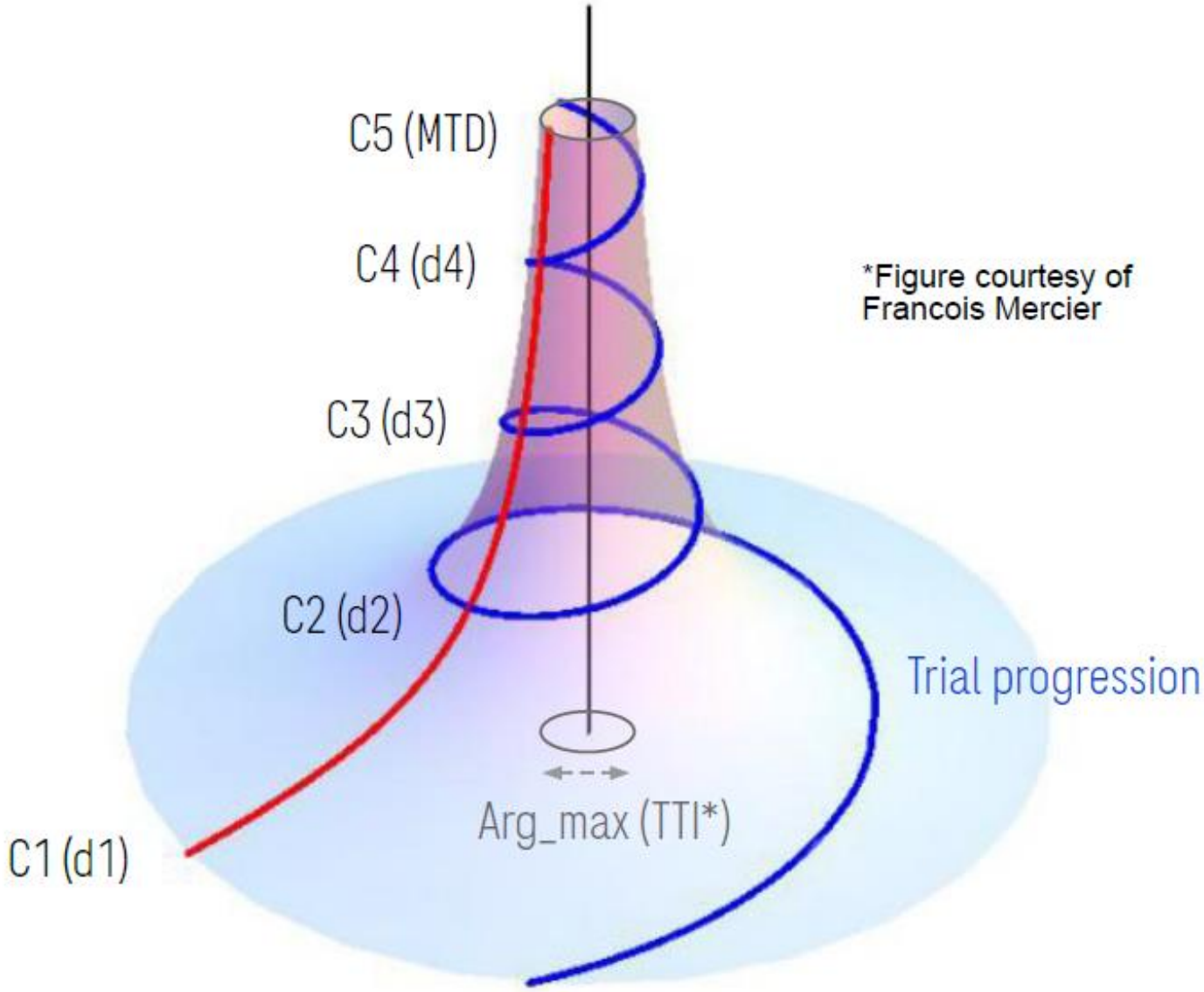


*Figure courtesy of Francois Mercier

Possible ICEs and Recommended Strategies

- Treatment discontinuation before the end of the DLT assessment period for reasons other than toxicity
  - Disease progression

Hypothetical or While on-treatment

- Dose modification
  1. Type 1: intentional, or unintentional but frequent
     - Dose reduction to prevent a DLT

While on-treatment

  2. Type 2: unintentional and/or sparse
     - Inability to receive the drug due to COVID closures

Treatment policy

Strategies: Discontinuation Due to Non-Toxicity Reasons

Strategy	*Targeted clinical question*	*Estimator consideration*
Treatment policy	What is the probability of DLT irrespective of the participant discontinuing treatment? *The targeted question is not clinically relevant*	Not applicable
Composite	What is the probability of DLT or treatment discontinuation? *The targeted question is not clinically relevant*	Not applicable
Hypothetical	What would be the probability of DLT, had treatment discontinuation not occurred? *The targeted question is clinically relevant*	Possible, e.g. using TITE-BOIN or DA-CRM
While on-treatment	What would be the probability of DLT, before treatment discontinuation occurs? *The targeted question is clinically relevant*	Possible, e.g. using TITE-CRM or TITE-EWOC
Principal stratum	What would be the probability of DLT, in the strata of participants who would not experience treatment discontinuation? *The targeted question is clinically relevant*	Difficult

Your Turn

- QUESTIONS?
  - The chair will moderate questions to the Panel

End

- THANK YOU!