Estimands and Safety in Oncology Clinical Trials

Jonathan Siegel, Bayer

Janet Wittes, Wittes LLC

Brenda Crowe, Lilly

Pedro Lopez-Romero, Novartis

Natalia Kan-Dobrosky, Abbvie

Science For A Better Life

The Pharmaceutical Industry Working Group on Estimands in Oncology

& PHUSE

Safety Task Force

Joint Statistical Meetings

August 6, 2023

Disclaimer

*The views presented in this presentation are **solely those of the authors,** and do not necessarily represent the views of the Pharmaceutical Industry Working Group on Estimands in Oncology, PHUSE, the authors' employers, or any other organization.*

Outline

- Purpose
- Safety Task Force
- The Cinderella of biostatistics?
- Safety issues
- Oncology-specific issues
- Problems with while-on-treatment strategies
- Case Study 1: Schlumberger et al., 2019
- Case Study 2: Weber et al., 2009
- Case study discussion
- Summary and conclusions

Purpose

- This presentation assumes general familiarity with the ICH E9 (R1) guidance.
- The purpose of our task force and this presentation is not to reformulate existing activities into new terminology.
- Rather, it is to attempt to apply the estimands framework to think more carefully about safety and formulate better and more fit-for-purpose (Deming 1986) clinical questions that can be answered more reliably.
- This, as we've found, is a very hard task.

Safety Task Force

- The Safety Task Force is a joint venture of the Pharmaceutical Industry Working Group on Estimands in Oncology (EIO) and PHUSE.
- EIO is a group of 97 industry statisticians and clinicians (mostly statisticians) from the US, Europe, and Asia.
  - From 47 companies.
  - Active regulatory participants.
  - Organized in 2018 to discuss the implications of ICH E9 (R1).
  - Affiliated with ASA-BIOP as the Estimands in Oncology Scientific Working Group
  - Also affiliated with EFSPI as a Scientific Interest Group.
- PHUSE is a global healthcare data science community organized in 2004 to provide a collaborative voice for data science professionals in the pharmaceutical industry.

Safety Task Force Progress

- The Safety Task Force was formed in March 2022 out of a recognition by EOI leadership that safety had not gotten sufficient attention.
  - None of the EIO members at that time were safety experts.
  - EIO partnered with PHUSE to create the task force and bring in analytics and safety experts.
- We completed a literature review including lectures by paper authors early this year, and began working on papers.
  - We have made less progress than expected.
- Safety, particularly in oncology, is hard.
  - Having safety experts learn about estimands and estimands experts learn about safety has involved a steep learning curve for both.
- Accordingly, this presentation of our preliminary work presents more questions and issues than answers.

The Cinderella of Biostatistics?

- On July 23, 2020, EFSPI had a session entitled “Safety Analyses: The Cinderella of Biostatistics?”
- the ICH E9 (R1) guidance refers to obtaining a clear description of the benefits *and risks* of a treatment. (Schuler, 2020)
- Yet safety rarely gets as much attention as efficacy in e.g. ASCO presentations. (Kubler, 2020).
- A great deal more is published in clinically-focused statistics journals on efficacy than on safety. (Kubler, 2020)
- Statisticians who rise to department heads or lead methodology groups in pharma companies are generally efficacy experts. (Kubler, 2020)
- Clinical trial safety reporting often consists of standardized incidences of AEs, labs, etc. -- perfunctory, routine, and descriptive.

The Cinderella of Biostatistics Cont.

- Assessments of causality tend to be clinically focused.
  - Clinical judgment attributing individual events in individual patients to treatment based on patient narrative.
  - Use of statistical methods in assessing causality at the clinical trial level remains limited.
- Little thought tends to be given to exactly what questions safety reporting and analyses are intended to answer.

Safety Issues:
Difference in posited causality

- Both efficacy and safety can involve the prevention of untoward events.
  - Efficacy events posited as effects of the underlying disease.
  - Safety events are posited as effects of the treatment.
- The difference in causality inverts conservativism.
  - Placebo is the quintessential inefficacious drug but the quintessential safe one.
  - Ignoring refusal of/withdrawal from treatment tends to reduce apparent efficacy, but increase apparent safety.
- Has an analogy to non-inferiority.
  - Both involve establishing an absence of a meaningful difference.
  - Both require care in distinguishing absence of evidence from evidence of absence.

Three kinds of causality

- Consideration of causality requires careful definition of what type of causality is involved.
- Safety evaluations may involve one of three kinds of causality.
  - Attribution of an individual safety event in an individual treatment to a specific drug regimen or drug.
  - Estimation of drug safety effect for a population (single-arm).
  - Comparison of safety across treatments.
- The 2012 FDA Safety Reporting Requirements moved responsibility for determining causality in expedited reporting from the investigator (per ICH E2A) to the sponsor.
  - Causality has nonetheless largely been determined by clinical judgment and descriptive statistical methods.
- The estimands framework potentially opens the door to considering more inferential statistical assessments of causality.

Safety Issues cont.

- ***Exploratory character.*** While specific safety objectives can be important parts of oncology trials and development programs, safety as a whole retains an exploratory character.
  - Previously unknown safety effects could be observed at any trial phase.
  - A clinical trial safety evaluation cannot be limited to focus on specific estimands, but must also look for whatever may be out there.
- ***Discrete clinic visits***. Most safety events require clinic visits to observe.
  - It is often infeasible to continue the physical exams, labs, etc. required for comprehensive safety much beyond treatment withdrawal.
    - More difficult than just limited efficacy assessments (e.g. imaging).
- ***Rare events***. Important safety events may be too infrequent to detect reliably within the scope of a clinical trial.

Oncology-Specific Issues

- Late-stage cancer has high morbidity and mortality.
  - Deaths are common events, and it can be difficult to distinguish death caused by treatment toxicity from death caused by the underlying disease.
- Cancer treatments are rarely curative. Most cancer-efficacy trials are designed to assess prolongation of progression and/or survival.
- Cancer treatments are often highly toxic.
  - Phase 1 studies are generally conducted in end-line patients rather than healthy volunteers.
  - High general toxicity can make a safer treatment more valuable, which should increase the importance of safety evaluation.

Oncology-Specific Issues cont.

- Randomization and blinding can be difficult.
  - Single-arm trials are not uncommon.
    - Experimental treatment may be ethically preferable to placebo if no effective alternative therapy.
  - Some treatment classes have signature side effects or other indicators enabling patients and investigators to de facto ascertain their assigned treatment.
- Subjective causality can be difficult and unreliable.
  - Lengthy time-to-event trials are common.
  - Immunotherapies and other recent therapy classes can have delayed and prolonged effects, decoupling the timing of safety events from treatment timing.
  - It is often unethical to wait to assign treatment long enough for effects of prior treatment to fully wash out.

Oncology-Specific Issues – Sample Sizes

- High morbidity and high toxicity tend to result in small sample sizes.
  - General tendency towards high toxicity limits ethical early-phase sample sizes.
  - High hazards of typical efficacy indicators like disease progression and death result in smaller late-phase trials than in some other therapeutic areas.
  - Multi-stage trials are common. Early termination for efficacy can result in even smaller sample sizes with which to assess safety.

A problem with while-on-treatment strategies

- Simple incidence reporting during treatment and to a time shortly afterwards, the most common approach used for reporting AEs, could be characterized as a while-on-treatment or while-at-risk strategy.
- As ICH E9 (R1) notes, “Particular care is required if the occurrence of the intercurrent event differs between the treatments being compared.”
- The basic problem is that such strategies treat both the intercurrent event and what happens afterwards as irrelevant.
  - This is often not a clinically appropriate assumption.
- As Hahn and Zhou (2023) note in a PRO context, “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 subdistribution hazard methods also have this bias.

A problem with while-on-treatment strategies (cont)

- The same would apply to other intercurrent events, like progression, that tend to result in ending assessments.
- Checking for superiority or non-inferiority on the relevant intercurrent event first, as an event in its own right, provides some protection from this bias. (Siegel, 2023)
- If a treatment provides a worse survival benefit, then any apparent benefit for lesser safety events may not matter as a practical consideration.
  - This may not be the case, however, for other intercurrent events leading to treatment withdrawal.
- In addition to potentially not accounting for intercurrent events in a clinically meaningful way, simple incidence reporting does not take time or exposure into account.

Another problem with while-on-treatment approaches

- Yang, Wittes, and Pitt (2018) provide a non-oncology example case where such reporting can be particularly problematic.
- They evaluated the SAVOR-TIMI 53 trial of saxagliptin vs. placebo in Type 2 diabetes mellites.
- The trial results had shown a mortality hazard ratio of 1.18 for the period up to 30 days past last treatment.
- Yang et al.'s reanalysis showed a higher mortality rate in placebo patients following treatment withdrawal, mortality not captured in the reported rates.

Yang et al. (2018) cont.

- As the authors explained,

- *"One likely explanation in this case is that nonadherence to placebo was a marker of having experienced a life-threatening adverse event; once such an event occurred, the participant was likely to stop randomized treatment and would have a higher mortality rate than those remaining on their randomized treatment during the same time period."*

- That is, the authors hypothesized that placebo patients tended to seek alternative therapy following a life-threatening safety event, with fatality only occurring after treatment withdrawal and hence not counting in the while-on-treatment analysis.
  - Omitting these events resulted in an incorrect conclusion.

Case Study: Schlumberger et al. (2015)

- Randomized, double-blind Phase III study of lenvatinib versus placebo in radioiodine-refractory thyroid cancer.
  - Primary endpoint PFS.
  - 392 patients randomized 2:1.
- Patients were unblinded, and placebo patients were permitted to crossover to open-label Lenvatinib, following disease progression.
- Successful trial.
  - PFS HR 0.21, 95% CI [0.14, 0.31], $p < 0.001$.
  - OS HR 0.73, p=0.10.
  - Median treatment duration on blinded study treatment was 13.8 months for Lenvatinib vs. 3.9 months for placebo.

Schlumberger et al. (2015) Cont.

- Published study safety analysis evaluated standard AE incidence during the blinded study treatment period.

	Lenvatinib	Placebo	
Median treatment duration (mos)	13.8	3.9	
Grade 3+	75.9%	9.9%	
Grade 5	2.3%	O%	

- Levinatinib was duly approved based on the study results.
  - The study achieved its purpose from a pharma company point of view.
- Nonetheless, the study design issues may have inhibited a proper understanding of lenvatinib safety.

Schlumberger (2015) Considerations

- Safety of open-label Lenvatinib treatment was not considered.
  - A “while during blinded treatment” strategy to address crossover prevented considering the complete safety profile of the experimental drug.
- The crossover design and analysis approach permitted evaluating only relatively short-term effects.
- As the example illustrates, the use of simple incidence may negatively bias the apparent safety of a more efficacious, longer-duration treatment.

Schlumberger (2015) Considerations cont.

- When AEs involve hazards over exposure and/or time, a longer duration of treatment and corresponding treatment exposure caused by a more efficacious treatment, as observed in this study, tends to increase the observed incidence of safety events.
- Can be true even when the hazards per unit time or exposure are the same or lower.

Case 2:Weber et al. (2009)

- Randomized, double-blind, placebo-controlled Phase II safety study comparing the tolerability and efficacy of ipilimumab with or without prophylactic budesonide in patients with unresectable Stage III or IV melanoma.
  - Patients were given ipilimumab with either prophylactic budesonide or placebo.
  - 115 patients randomized 1:1.
  - Primary endpoint was incidence of Grade 2+ diarrhea
- Study was unsuccessful, with similar rates of primary endpoint:
  - Ipilimumab+budesonide: 33% (95% CI: 21-46)
  - Ipilimumab+placebo: 35% (95% CI: 23-49)

Weber et al. (2009) Cont.

- A potential issue is effect over time vs. worst incidence.
  - Might budesonide have reduced the duration of moderate+ diarrhea even if it didn't meaningfully affect its incidence?
- "Duration" and "number of events" concepts might be difficult to implement consistently.
  - An "event" was not defined in the publication.
  - Diarrhea tends to be episodic. Is each episode an individual AE or is it one AE with an extended duration?
  - Different investigators might have different opinions based on subjective judgment if not carefully defined.
  - Without careful definition, not clear what the statement "no patient experienced more than 2 events" meant.

Weber et al. (2009) Cont.

- It might be useful to consider the effect of time.
- Alternative approaches might be:
  - Time to first incidence.
  - Time/exposure-adjusted diarrhea rate.
  - Diarrhea-free days
- It might be useful to consider the effect of intercurrent events resulting in informative loss to follow-up, such as death or treatment withdrawal due to other AEs.
- Example approaches might be:
  - Composite strategy (e.g. diarrhea-free survival).
  - CIF (death as competing risk).

Example discussion

- The Schlumberger (2015) study aimed to establish efficacy in cases where existing therapy provided a poor prognosis.
- The strong efficacy benefit likely outweighed any weaknesses in the safety evaluation.
  - If a treatment provides a strong survival benefit in a high-mortality disease, a clearer understanding of the safety risks may well not affect approvability.
  - A crude evaluation of safety may be sufficient so far as "fitness for purpose" (Deming 1986) is concerned.

Example Discussion cont.

- In the Weber (2009), study, however, the purpose of the study was to determine whether a concomitant treatment provided a safety advantage.
- Where safety is especially relevant, outside the context of treatments with overwhelming survival efficacy, time-dependence and the effect of duration of treatment may be particularly important.

Summary and conclusions

- Current standard approach to safety reporting could be characterized as a while-on-treatment (while at risk) strategy for intercurrent events ending treatment (death, progression, other toxicity).
- It is significantly flawed when treatment durations diverge or when time to treatment-ending events like progression or death differs between the treatments.
  - Approach is biased against the longer-duration, longer-survival treatment.
- When a treatment provides a strong efficacy advantage, especially a survival advantage, biases introduced in the safety reporting may not make enough of a difference to affect overall approvability despite potentially affecting clinical interpretation.

Summary and conclusions cont.

- In cases where safety is particularly important, however, clearer and more careful reporting of safety that takes duration and the effects of intercurrent events into account is essential.
  - Safety trials.
  - Specific safety objectives, especially when following for known safety issues.
- In cases where a strong difference in treatment duration is plausible, a method taking exposure into account is suggested, at least as an alternative analysis.
  - Safety hazards are rarely constant, so methods assuming constant hazards can have their own biases.
- As Yang et al. suggest, safety analyses should look for patterns of deaths or other key events beyond treatment withdrawal.

Summary and conclusions cont.

- The estimands guidance helps provide a framework for more clearly understanding and addressing how these can lead to more valid and reliable assessments of safety.
- Our preliminary work so far has resulted more in raising questions than providing answers.
- That said, raising important questions and giving careful thought to them is a critical part of what the estimands framework is designed to achieve.
- Our working assumption has been that safety, characterized as the “Cinderella of biostatistics,” could particularly benefit from this process.
- We continue to think this warranted.

Thank you!

References

- Hofner, B. Safety analyses: The Cinderella of biostatistics? Regulatory perspective. Presented at EFSPI 11th Statistical Leaders Meeting 07.23.2020. https://www.efspi.org/Documents/Events/Events%202020/Stat%20Leaders%20Meetings/02_EFSPI%202020_02_SafetyAnalyses_Regulatory.pdf
- Kubler, J. Safety analyses: The Cinderella of biostatistics? Industry perspective. Presented at EFSPI 11th Statistical Leaders Meeting 07.23.2020. https://www.efspi.org/Documents/Events/Events%202020/Stat%20Leaders%20Meetings/02_EFSPI%202020_03_SafetyAnalyses_Industry.pdf
- Schuler, A. Safety analyses: The Cinderella of biostatistics? Industry perspective. Impulse presentaiton. Presented at EFSPI 11th Statistical Leaders Meeting 07.23.2020. https://www.efspi.org/Documents/Events/Events%202020/Stat%20Leaders%20Meetings/02_EFSPI%202020_01_SafetyAnalyses_ImpulsePresentation%20(002).pdf
- Deming, WE. *Out of the Crisis*. MIT Press (1986)
- United States Food and Drug Administration. Safety Reporting Requirements for INDs (Investigational New Drug Applications) and BA/BE (Bioavailability/Bioequivalence) Studies (2012).
- United States Food and Drug Adm inistration. E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials (2021)
- Hahn, S and Zhou, XH. Defining estimands in clinical trials: A unified procedure. *Stat in Med* 42: 1869-87 (2023)
- Siegel, J. et al. Time-to-event estimands and loss to follow-up in oncology in light of the estimands framework.Submitted to *Pharmaceutical Statistics*. https://arxiv.org/abs/2203.01781
- Yang, F., Wittes, J., and Pitt, B. Beware of on-treatment safety analyses. *Clin Trials* 16:63-70 (2019).
- Schlumberger, M. et al. Lenvatinib versus Placebo in RadioiodineRefractory Thyroid Cancer. N Engl J Med 372:621-30 (2015)
- Weber, J., et al. A randomized, double-blind, placebo-controlled, Phase II study comparing the tolerability and efficacy of ipilimumab administered with or without prophylactic budesonide in patients with unresectable Stage III or IV melanoma. *Clin Cancer Res*. 15:5591-8 (2009)