The Estimand Academy for Trial Teams
"Brining estimands to life through real case studies"
Webinar 2: Estimands in Oncology – How and Why

Most frequently asked questions during the webinar "Estimands in Oncology - How and Why"

Disclaimer: *The answers represent the views of the presenters from the clinical engagement sub-team of the Estimands in Oncology working group (www.oncoestimand.org). They do not represent official guidance or policy of industry or authorities.*

General Questions

Question: Is the estimand framework only to be applied for registrational studies (Phase 2 and 3) or also for Phase 1 studies?

Answer: The framework is useful to describe every endpoint, irrespective of phase, TA, etc.
The addendum says at least every endpoint "pertaining to regulatory decision making" should use the framework, but it is useful for any endpoint (efficacy, safety, PRO, PK/PD, single-arm, RWD, etc.) where you want to describe an effect.

Question: From the five strategies which should be used?

Answer: The strategy needs to be aligned to your clinical trial objective, there is no "right" or "wrong" or "better" strategy. A clear scientific objective will help in selecting appropriate estimands and strategies. It is important to understand the clinical question is being asked in light of the purpose for which the results will be used. Teams should also assess the clinical relevance and feasibility of estimating the desired estimand in the proposed setting. A treatment policy strategy, for example, require systematic follow-up through and beyond the relevant intercurrent event, and this level of follow-up is not always feasible in every study context. We advise that the team consider different strategies, as there is often no perfect strategy or there are limitations with regard to data collection or visit schedules that limit the value of a particular strategy. Ultimately, one needs to define one primary strategy for every intercurrent event. It can happen that all ICEs are dealt with using the same primary strategy, but a mix of strategies is perfectly fine. Alignment to the clinical trial objective is the key.

Question: Many protocols are still phrased in terms of primary and secondary endpoints. Should they be?

Answer: Yes, but extend endpoint description to estimand description (of which endpoint / variable is just one of five attributes).

Question: Per Estimand principle, some protocol deviation items will be considered intercurrent events to be handled if interested; so no more per protocol population needed?

Answer: At least a naive per protocol analysis is not encouraged according to the estimand framework.

Question: Is it Mandatory to define estimand in protocol or can we provide the detail in SAP?

Answer: Estimand discussion should happen during protocol development or even earlier, it helps to define the treatment effect we want to estimate more precisely, and then study design, data collection and analysis plan shall align with the Estimand you defined. So, we would highly recommend you to define the estimand in protocol. For those studies with already finalized protocol, we may then go with the alternative to have that in the SAP. A critical reason for describing the estimand in the protocol is that it needs to be aligned with the visit schedule, and the consequences should be considered when the schedule is being designed. For example, ending clinic visits impacts all estimands that depend on clinic visits and may alter both the strategies that can be appropriately used for those estimands and their usability and interpretation. Sometimes when to end visits may need adjustment taking into account considering all relevant estimands. The study should be designed with knowledge of these consequences to create the best feasible study doing under the circumstances.

Question: How do you suggest writing the estimand, in one sentence with details either below it or in a separate section? Or do you suggest another way of presenting?

Answer: The key is that all five attributes of the estimand are described in detail. The checkmate case study used a verbal presentation. A tabular formulation of estimands with repeating attributes that do not change is an efficient approach that is in use as well.

Question: How often is the principal stratum strategy used? What is the regulatory acceptability of it given the often strong assumptions that are needed for analysis?

Answer: There is some precedence for regulatory acceptance. Please also see:
Bornkamp et al. Principal Stratum Strategy: Potential Role in Drug Development (2021). Pharmaceutical Statistics https://doi.org/10.1002/pst.2104

Question: How does the estimand framework combine with multiple testing issues?

Answer: The description of estimands is generally unrelated to multiple testing. In general, description of estimands is useful for any endpoint (addendum says for "endpoints pertaining to regulatory decision making", but framework is useful for any endpoint), irrespective of how this endpoint is

used within a multiple testing framework. Each comparison is formulated separately. That said, one thing the estimand framework can sometimes help with is to clarify dependencies among the estimands. Whether estimands are statistically independent or not can depend on study design and choice of strategy. These dependencies can sometimes affect the multiple testing strategy.

Question: Main and sensitivity analyses are performed on the same estimand, and changing a strategy will generally change an estimand. Changing a strategy is not a part of sensitivity analyses. Sensitivity analyses are to change assumptions used in statistical analyses such as missing at random (MAR) and not missing at random (NMAR). Is this understanding correct?

Answer: Yes, this is also our understanding. It is generally recommended to perform a sensitivity analysis to address each of the key assumptions made by the strategy and analysis method used.

Question: Are special considerations needed to use the estimand framework in single-arm trials?

Answer: The framework is useful to describe every endpoint, irrespective of phase, TA, etc.

In some cases, a different estimand might be appropriate for a single-arm trial than for a comparative one. For example, in a comparative trial a treatment policy strategy can provide an appropriate scientific assessment because randomization permits comparing a treatment policy initiated with the study drug from a treatment policy initiated with a comparator. But in a single-arm trial, it can be more difficult to determine whether the overall efficacy observed in the trial was due to the study treatment or if similar results would have occurred with standard of care plus subsequent therapy. For this reason, the treatment policy strategy generally requires more caution in interpretation in a single-arm context than in a randomized comparative context, and a different strategy may sometimes be appropriate.

There are also some unique questions to single-arm trials that need further discussion. For example, an intercurrent event is defined as occurring after treatment initiation. In randomized trials, in most cases the treatment initiation is the moment of randomization. How would treatment initiation be defined in single-arm trials? If a patient withdraws consent before receiving the first dose of the treatment, would that be an intercurrent event, or would it be considered as occurred before treatment initiation? The decision how to define these will be connected to the scientific objective of your single-arm trial.

That said, the estimands framework sometimes imposes additional scientific rigor on clinical trials that limits what development teams can do. Other aspects of regulatory scientific rigor, such as full family-wise error control, performing interim analyses only at prespecified times, etc. are commonly relaxed in early phase studies. A development program which seeks to maximize flexibility in early phase may also not wish to limit itself by some scientific-rigor elements of the estimands framework. In general, study teams should only adapt the level of scientific rigor they are willing to abide by.

The Oncology estimand working group has established an early development estimand nexus (EDEN) task force that specifically aims at bringing together statisticians from industry, regulators, and academia to ensure common understanding and consistent definitions for key estimands in early oncology and share experiences, intercurrent events, and the used sensitivity analyses. This task force will also address single-arm trials.

Specific Questions on the Case Studies

Question: Can the panel please elaborate on the example used in the quiz where one of the options was "censor at time of event".
I'm not sure I understand why this is not a sensible option from a stats point of view?

Answer: Censoring at the start of subsequent therapy assumes that changing therapy has no effect on patients' future risks (in statistician-speak, subsequent therapy is not "informative"). This assumption forms a hypothetical scenario and may also be a very dubious one clinically. Patients generally stop study treatment and get subsequent therapy when they think the study treatment is not working for them and/or a different therapy could do better. This often means they have gotten worse and are likely to be at higher risk than patients able to continue on the treatment. The new therapy itself might also change their risks. Assuming that patients receiving new therapy have the same risks as patients who continue on study therapy rarely makes clinical sense and likewise may introduce informative censoring which can lead to an inflated type-I error, a biased hazard ratio estimate and may impact statistical power. If used, in view of the estimand framework a hypothetical strategy is applied. In this hypothetical strategy the relevant scientific question is what would have happened if the relevant event had not occurred. A key distinction to understand the answer to this quiz question is that censoring is an estimation method and does not "live" on the level of estimand (defining what is to be estimated), but on the level of the estimator (defining how it should be estimated). So, whereas censoring at the time of the event might be an appropriate way to estimate your estimate for a particular estimand, it implies a hypothetical strategy for the estimand.

Question: Any other way to perform the hypothetical strategy for time to event endpoints, other than censoring at the intercurrent event?

Answer: Yes, these exist, e.g., Rank Preserving Structural Failure Time (RPSFT) or inverse-probability-of-censoring weighting (IPCW). These methods don't require assuming non-informativity, i.e., that the intercurrent event does not change future risks for the main event. But they make other, sometimes even stronger assumptions.

Question: For the population summary measure, when we have a time to event endpoint (e.g. PFS, OS), how would you incorporate the log-rank test in the estimand?

Answer: The test statistic itself is not one of the five components of the estimand. With a time-to-event endpoint there are numerous test statistics we could look at (e.g., log-rank, weighted log-rank, restricted mean survival time [RMST], etc.).

Question: What is the value of the hypothetical strategy in a clinical trial when we know that the intercurrent event can happen in real life. Like starting new anti-neoplastic therapy.

Answer: One case where a hypothetical strategy can be helpful is where the clinical trial environment is different from the environment we really want to study. For example, in some open-label trials, a high percentage of patients randomized to placebo immediately withdrew from study treatment and enrolled in a different clinical trial, in some cases testing a different experimental drug in the same class as the study treatment. This behavior would not happen in clinical practice. Patients with serious diseases don't withdraw from treatment without ever trying it. And patients discontinuing failed therapy rarely take therapy in the same class. It happens because a clinical trial environment, especially its use of placebo, is artificial, and induces behavior that isn't realistic. The artificial, trial-environment-only behavior may be so pervasive as to confound the results of a treatment policy strategy. When this occurs, the clinical question posed by a hypothetical strategy, "what would have happened if the artificial treatment-switching behavior hadn't occurred?" may be the right question. It may actually be more realistic and clinically meaningful than the question posed by a treatment policy strategy, "what will happen if treatment behavior (the treatment policy) in the clinical practice environment is similar to the treatment policy in the trial?" In these cases, where the clinical trial environment is counterfactual to real-world clinical practice, a counterfactual hypothetical strategy should be considered.