Statistics in Biopharmaceutical Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/usbr20

What can be Achieved with the Estimand Framework?

Susan Mayo & Yongman Kim

To cite this article: Susan Mayo & Yongman Kim (2023): What can be Achieved with the Estimand Framework?, Statistics in Biopharmaceutical Research, DOI: 10.1080/19466315.2023.2173645

To link to this article: https://doi.org/10.1080/19466315.2023.2173645

What can be Achieved with the Estimand Framework?

Susan Mayo[a]* and Yongman Kim[b]

[a]*Senior Mathematical Statistician, Division III, Office of Biostatistics, Office of Translational Science, Center for Drug Evaluation, Food and Drug Administration, Silver Spring, MD, USA*

[b]*Lead Mathematical Statistician, Division III, Office of Biostatistics, Office of Translational Science, Center for Drug Evaluation, Food and Drug Administration, Silver Spring, MD, USA*

*Correspondence details for the corresponding author: susan.mayo@fda.hhs.gov

Abstract

The ICH E9(R1) guidance on estimands is a key tool for the creation and review of protocol design and analysis planning, for both industry and regulatory statisticians. The framework has been described as useful for improving study design, intercurrent event handling, data collection, analysis, and interpretation to align the estimand with the primary clinical question to add clarity and precision to support regulatory decision-making. In this paper, we describe our experience as regulatory statisticians in review of Investigational New Drug protocols and statistical analysis plans, with an emphasis on trials used to support substantial evidence of effectiveness in New Drug Applications and Biologic License Applications. Our intent is to describe our experience with this powerful and effective framework tool, to align the clinical trial's primary objective with its analysis outcomes and interpretation.

Keywords: Estimand; ICH E9(R1) Guidance; Intercurrent event; Missing data; Sensitivity analysis

1. Introduction

An estimand is described in the Food and Drug Administration E9(R1) Statistical Principles for Clinical Trials: Addendum as "a precise description of the treatment effect reflecting the clinical question posed by the trial objective. It summarizes at a population-level what the outcomes would be in the same patients under different treatment conditions being compared."

The estimand structure is a simple-to-use and highly valuable means for aligning the estimand attributes,[1] intercurrent event (IE) strategies, estimator and sensitivity analyses back to the clinical question (Figure 1). Motivated by Bell, et al. (2021), once the estimand attributes are identified, one can more precisely construct the clinical question posed by the trial objective, particularly when proposed objectives or estimand attributes are ambiguous. Together, the primary objective, estimand attributes, strategies for intercurrent events (IEs) and missing data are the basis for appropriate consideration of the primary estimator and sensitivity analyses. Use of the estimand framework in this way avoids ambiguous interpretation of trial results which can lead to regulatory hurdles. As described by Ratitch, et al (2020), the DIA Estimands and Missing Data Working Group recommends a related approach. When providing advice to sponsors on study design, trial conduct, data collection and analysis we authors focus on the estimand framework. We introduce these concepts early in development so they can be implemented at the right time. If the drug is deemed safe and effective, the estimand framework can also inform that drug's labeling.

[1] Attributes are the treatment condition of interest, population of patients targeted by the clinical question, variable obtained from each patient to address the clinical question, population-level summary for the variable, and intercurrent events.

2. Estimand Development History

Statistical principles for clinical trials are well delineated in the ICH E9 guidance published in 1998. It emphasizes intent-to-treat (ITT) principles and the associated treatment policy effect as a target of estimation. However, it was felt that further guidance on handling missing data was necessary to estimate the treatment effect reliably, especially in the presence of large amounts of missing data. Therefore, FDA commissioned the National Academy of Science (NAS) for advice on handling missing data in clinical trials, and subsequently NAS looked to its National Research Council with over 30 members from academia, industry, and regulatory agencies to make recommendations. In 2010, the council issued a report entitled "Prevention and Treatment of Missing Data in Clinical Trials." One of the major recommendations was to distinguish "analysis dropout" from "study dropout" and to continue to collect efficacy and safety data even after discontinuation of treatment to minimize missing data. In addition, the report recommended to define an estimand aligned with the trial objective in the protocol. Adoption of this recommendation by trial sponsors led to the R1 addendum to ICH E9 guidance to add clarity on the estimands, including an introduction of "intercurrent events" (IEs) as an essential component, and to provide example strategies to account for such events. The guidance envisions different strategies to address various IEs in addition to treatment policy strategy strongly advocated in ICH E9 guidance. The ICH E9(R1) guidance was finalized in 2019 and adopted as FDA guidance in 2021.

3. How statistical reviews benefit from ICH E9(R1)

Traditionally trial objectives are defined in protocols at a very high level such as demonstrating efficacy as primary and safety/tolerability as secondary, but this definition may not be helpful because it is not specific enough to understand what is to

be estimated to support the primary trial objective. When regulatory statisticians review a protocol or statistical analysis plan (SAP), often we need to search many sections of the protocol for elements scattered throughout to understand the main clinical objective of interest. A certain amount of "educated" guesswork is made for any missing elements. Reviews of some protocols lead us to the conclusion that what was estimated is not aligned with the stated clinical objective.

Providing a clear estimand in the protocol can address these uncertainties. Identification of the estimand attributes in the protocol, prior to initiation of the trial, ensures these attributes are considered. Use of the estimand attributes and strategies for both IEs and missing data can lead to more appropriate study design, data collection, study conduct, and important features of data analysis by considering these elements at the right time for implementation. In the protocol or SAP, further analysis details may be pre-specified prior to unblinding. Regardless of which document captures the estimand details, of primary importance is that the IE strategy appropriately reflects the clinical question of interest, and that the primary trial objective and its affiliated estimand attributes are aligned with regulatory decision-making. Estimand choices used to support regulatory decision-making typically inform the label (see Table 1).

Without an estimand and awareness of a clear primary clinical objective designed to answer the clinical question for regulatory approval, sponsor statisticians might tend to propose a statistical model to generate some "statistically reasonable" estimates and related inferences. Therefore, how the IEs and missing data were addressed in the statistical analysis could be based on a statistician's experience and discretion based on habitual choices, possibly with their clinician colleague's tacit agreement and without consideration of the impact on the trial's interpretation. Interpretation of treatment effects defined solely by statisticians was usually not

transparent in the presence of premature treatment discontinuation, rescue medications and procedures, changes in background therapy, use of prohibited medication, non-ascertainable patient outcomes, and non-random patterns of IEs and/or missing data.

Our experience to date is that the estimand framework in E9(R1) can resolve most of such issues by considering the clinical question to support regulatory approval at the trial design stage with detailed plans on how to address IEs such as those listed above. Furthermore, clearly defining the other four attributes of the estimand helps with alignment of the other attributes, such as target population and summary measure with the primary clinical objective.

Certain traditional approaches of using statistical methods to address IEs should be considered, as appropriate, to answer the clinical question to support regulatory approval. E9(R1) clearly defines "missing" data separate from IEs and recommends missing data to be addressed "statistically," but IEs to be addressed "clinically."

4. How we use estimands in our reviews

E9(R1) narrows the gap between regulatory and sponsor statisticians on what is to be estimated and how it is estimated. This guidance encourages regulators to engage in multidisciplinary discussions prior to communication with sponsors. These experiences accumulate to reflect the evolution and improvements in estimand definitions across various therapeutic areas and promote transparency in regulatory guidance with respect to precedence and consistency.

In the IND protocol (and SAP) statistical reviews, we include the sponsor's description of the estimand, estimator and sensitivity analysis. In addition to our statistical reviews, clinical reviews may address whether the protocol adequately specifies and handles IEs. After documenting what estimand information the sponsor noted in the protocol or SAP, we provide our perspective on appropriateness.

Typically, the disciplines confer to form their recommendations to sponsors. Our experience has been that as statistical and clinical reviewers become more familiar with estimand terminology and usage, statisticians have become more aware of the nuances of clinical practice in each disease, how that may affect clinical trial interpretation and thus, the IE strategies. Our experience in using estimands in our reviews has afforded us greater opportunity for consistency and for shifting important conversations to the right time: during the IND reviews of protocols for confirmatory trials. Sponsors then have this feedback for inclusion in their NDAs and BLAs in support of substantial evidence of effectiveness (FDA, 2019).

5. Nasal polyps example and how it was helpful for reviewers

Our initial use of estimands and ICH E9(R1) was based on discussion with clinical review colleagues in the Division of Pulmonology, Allergy and Critical Care. We agreed to use the estimand framework as a means of offering consistency across drug programs for commonly reviewed indications. We considered how the five attributes of the estimand and the strategies chosen to address post-randomization IEs aligned with the clinical question(s) to support regulatory approval, which is influenced by prior decisions in BLA, NDA, and IND reviews and labeling.

The first indication we focused on was chronic rhinosinusitis with nasal polyps (CRSwNP), beginning with the review and approval of the first biologic for this indication. In protocol and SAP reviews, and in response to sponsors' estimand questions, we included our view of the estimand attributes, IEs and appropriate estimand strategies for each of the IEs. We also described the estimator and sensitivity analyses that aligned with this estimand and placed it in the statistical issues section of our BLA reviews. Our views on the CRSwNP estimand evolved over time as more INDs, NDAs, and BLAs were reviewed, resulting in the current views for this

indication. Sponsors provided thought-provoking perspectives about estimands, estimators and sensitivity analyses in their protocols, SAPs and clinical study reports (CSRs). These perspectives were incorporated into what we considered to be the appropriate strategies for IEs and the relative importance of statistical properties of the estimator. We often found the ICH E9(R1) guidance helpful when considering sponsors' estimand and estimator descriptions. Table 1 is an example, derived from the recent FDA chronic rhinosinusitis with nasal polyp (CRSwNP) guidance (2021), of how the estimand attributes align with the clinical question and with the estimator and sensitivity analysis.

***

6. Estimand strategy is tied to the intercurrent event, and its practical importance

Appropriate choice of IE strategy is affected by clinical practice or the nature of the medication under study. For example, whether a concomitant medication IE should be considered a treatment failure (i.e., with use of a composite strategy as outlined in E9(R1)) may depend on a variety of factors such as the type of study treatment and the timing of when the concomitant medication was administered. For example, a systemic study medication would be less impacted by the IE of a short burst of oral corticosteroids during the first weeks in a year-long trial than would a study medication delivering a localized study treatment in a shorter duration trial.

Clinical context is crucial for interpreting the impact of various IEs on the real-world situations for patients. For example, determining the appropriate IE strategy for administration of a concomitant medication IE and whether its impact is considered part of the standard of care treatment regimen (thereby indicating a treatment policy strategy), or it indicates treatment failure of the study treatment (thereby indicating a

composite variable strategy) cannot be accurately determined unless medical knowledge of the disease and current treatment options are well understood. In another clinical setting a different strategy may be warranted.

7. Aligning trial objectives, design, planning, and analysis using the estimand framework

As noted in the introduction, it is easy to see that once the estimand attributes are identified, one can more precisely refine the clinical question (Figure 1). This has been helpful for comparing the sponsor's clinical question to one informed by the review team's construction of the estimand attributes in the sponsor's protocol. The estimand framework provides a mechanism for clearer communication of elements needed for a more precise clinical question and trial objective. It follows that a more precise primary clinical question based on attributes that also create alignment with the estimator, including use of appropriate IE strategy and handling of missing data, will improve interpretability of clinical trial data.

More precisely defining the clinical question and corresponding estimand has led to some challenges and differences in perspectives. For example, it is uncommon that we have found a hypothetical strategy to be the most relevant one for the primary clinical question used in regulatory decision-making (with exceptions, e.g., a pandemic leading to temporary site closures that are not expected to occur in general clinical practice in the future).

Further, we agree with Ionan, et al, (2022) that while using creative names for estimands, such as "efficacy estimand" or "attributable estimand" may be acceptable for ease of communication in some cases, the utility of such named estimands will be limited without corresponding sufficient clarity based on elements of the ICH E9(R1) when they do not relate precisely to the primary clinical question to support regulatory

approval. For the efficacy estimand (which identifies analysis of only data from the time when subjects were taking study medication to address a hypothetical scenario where such subjects continued to receive assigned treatment), when it is proposed as the primary estimand, this suggests the primary clinical question to support regulatory approval need not consider the data after patients discontinue study medication. While this may be a reasonable question for a sponsor to ask, from a regulatory decision-making perspective, it does not address how effective the drug is on average at the levels of adherence achieved in the trial (and expected in clinical practice), which is an important public health consideration. Furthermore, the approach relies on strong and unverifiable assumptions about the hypothetical outcomes if patients who discontinued treatment had remained on treatment.

With the estimand framework in mind, an alignment is naturally expected among the population summary measure and its point estimate, CI estimate and hypothesis testing for regulatory decision making. If a clinically interpretable solution cannot be found with a given estimand or estimator, an iterative process might be needed for redefining as described in Figure 1.

From the perspective of global implementation, although frequently aligned, agencies may have different views on how IE strategies and missing data should be handled, appropriateness of an endpoint as the primary endpoint, or differing views on appropriateness of statistical methodologies. In some diseases, the primary endpoints and statistical methodologies used to study them have attained global common practice and agency cross-alignment. One may anticipate the estimand attributes and choice of IE strategy in these cases will become common practice as well.

8. Conclusions

The estimand framework helps to achieve the goal of providing precise IND feedback for sponsors by (1) providing clear feedback that is consistent for each indication to enable sponsors to file submissions that meet both sponsor and regulatory expectations, (2) narrowing the gap on target of estimation in connection with IEs between sponsor and regulator, and (3) providing transparency in handling missing data by distinguishing IEs from missingness.

Acknowledgements: The authors would like to thank Miya Paterniti, MD, Rekha Jhamnani, MD, and Gregory Levin, PhD for their helpful review comments.

Disclosure Statement: This publication reflects the views of the authors and should not be construed to represent FDA's views or policies. The authors report there are no competing interests to declare.

References

Bell, J., Hamilton, A. Sailer, O., Voss, F. (2021), "The detailed clinical objectives approach to designing clinical trials and choosing estimands," Pharmaceutical Statistics. 2021;1–13.

Ionan, A., Paterniti, M., Mehrotra, D., et al, (2022), "Clinical and Statistical Perspectives on the ICH E9(R1) Estimand Framework Implementation," Statistics in Biopharmaceutical Research DOI: 10.1080/19466315.2022.2081601.

National Research Council. (2010), *"*The Prevention and Treatment of Missing Data in Clinical Trials*,"* Washington, DC: The National Academies Press. doi.org/10.17226/12955

Ratitch, B. et al., (2020), "Choosing Estimands in Clinical Trials: Putting the ICH E9(R1) Into Practice," Therapeutic Innovation & Regulatory Science 2020, 54(2) 324-341. doi: 10.1007/s43441-019-00061-x

U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Center for Biologics Evaluation, (September 1998), "ICH Guidance for Industry: E9 Statistical Principles for Clinical Trials," www.fda.gov/media/71336/download

U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Center for Biologics Evaluation, (December 2019), "Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products Guidance for Industry," www.fda.gov/media/133660/download

U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Center for Biologics Evaluation, (May 2021), "E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials Guidance for Industry," www.fda.gov/media/148473/download

U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Center for Biologics Evaluation, (December 2021), "Chronic Rhinosinusitis with Nasal Polyps: Developing Drugs for Treatment Guidance for Industry," www.fda.gov/media/154724/download

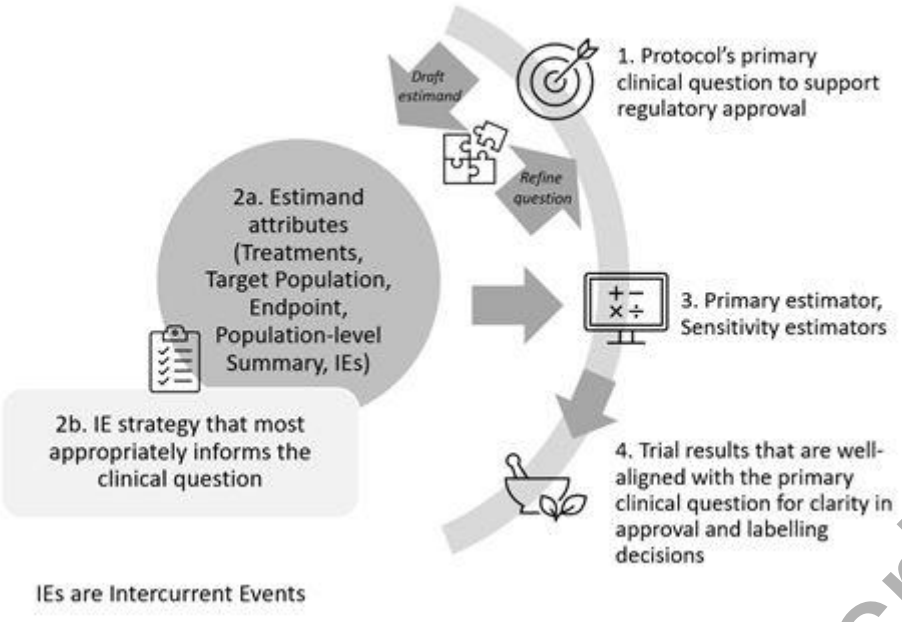

Figure 1. Steps used to identify and assess a protocol's or statistical analysis plan's estimand in our IND reviews. Steps 1 and 2: Estimand attributes with clinically appropriate intercurrent event handling are derived from, and inform, the primary clinical question. Step 3: The aligned objective, estimand attributes and IE strategies are used to identify an appropriate estimator and sensitivity analysis. Step 4: Clinical trial results are well-aligned with what is needed for the primary regulatory decision.

Table 1. An example estimand and clinical question to support regulatory approval, with considerations for estimator and sensitivity analysis based on the Chronic Rhinosinusitis with Nasal Polyps Indication guidance.

Estimand Attributes

Treatment conditions of interest: Active treatment of interest added to standard of care; placebo treatment added to standard of care

Population of interest: Adult patients with physician diagnosed chronic rhinosinusitis with nasal polyps

Co-primary endpoints: Change from baseline in co-primary endpoints, nasal polyp score (NPS) and nasal congestion score (NCS), with rescue surgical treatment considered as a type of treatment failure.

Population level summary: Difference in population means between active and placebo treatment groups at Week 52

How intercurrent events are reflected:

1. Rescue surgical treatment: patients who had nasal polyp surgery are considered to have had a treatment failure. Composite variable strategy indicates that the NPS or NCS measurement after the surgery is not meaningful, and the surgery event itself meaningfully describes the patient's outcome and is incorporated into the co-primary endpoints. (Composite variable strategy: intercurrent event is taken to be a component of the variable)

2. Rescue systemic corticosteroid treatment: (biologic products example) the clinical interpretation of the variables of interest (NPS or NCS) for the patients who did not have nasal polyp surgery is that patients are only temporarily impacted by systemic corticosteroid treatment and that understanding how effective the drug is on average on top of standard of care, including as-needed steroid use, is an important public health consideration. (Treatment policy strategy: the intercurrent event is considered part of the standard of care treatment regimen and does not impact the clinical interpretation of the variable of interest)

3. Treatment discontinuation: the clinical interpretation of the variables of interest (NPS or NCS) for the patients who did not receive rescue therapy, but discontinued treatment, is that understanding how effective the drug is on average at the levels of adherence achieved in the trial is an important public health consideration. (Treatment policy strategy)

Note: both co-primary endpoints must be statistically significant for the trial to be considered successful.

Detailed clinical objective to support regulatory decision-making, based on estimand attributes

The primary trial objective is to demonstrate superiority of active treatment of interest against placebo treatment added to standard of care in adult patients with physician-diagnosed chronic rhinosinusitis with nasal polyps.

Change from baseline in co-primary endpoints, nasal polyp score (NPS) and nasal congestion score (NCS) will be compared between active treatment of interest against placebo using difference in population means between active and placebo treatment groups at Week 52 as the population-level summary.

The treatment failure event of rescue surgical treatment may affect the treatment comparison and will be handled by considering patients who have surgery to have the worst possible scores for both NPS and NCS.

Recommendations for the design, conduct, and primary analyses

To reliably evaluate the estimand described above, the following are recommendations for the primary analyses of the co-primary endpoints:

- The analysis should be conducted in all randomized patients using all data for the 52-week duration.
- A regression-based approach is recommended to compare mean* changes at Week 52 between treatment groups.
- To improve the precision of treatment effect inference, recommend adjusting for prespecified prognostic baseline covariates (e.g., asthma/NSAID-ERD status, prior surgery history).
- Patients who have nasal polyp surgery should be considered to have had a treatment failure. One reasonable approach is to assign these subjects the worst possible score for the coprimary endpoints NCS and NPS.
- Data should be collected after treatment discontinuation or rescue medication use and all such data collected should be included in the analysis.
- Missing data should be addressed in the primary analysis using reasonable assumptions.
- Plans for sensitivity analyses for missing data assumptions and supplementary analyses should be pre-specified.

Considerations for sensitivity analysis

Explore the robustness to limitations in the data and deviations from the assumptions used in the statistical model for the main estimator through alternative missing-not-at-random assumptions (including analyses that comprehensively evaluate alternative assumptions such as a tipping point analysis).

Considerations for supplementary analysis

One analysis a sponsor may wish to explore in an alternative estimand is handling of systemic corticosteroid as an intercurrent event with a composite variable strategy.

* Rank-based analysis has been used to address treatment failure IEs by assigning lower ranks. But this approach is not aligned with the clinical question to estimate a clinically interpretable point estimate and measure of variability that is useful for healthcare providers and patients. Instead, a parametric analysis assigning the worst possible score to address treatment failure IEs is recommended to provide an estimate of the mean difference and its CI.