Understanding estimands

Nithya Jaideep Gogtay, Priya Ranganathan, Rakesh Aggarwal

Department of Clinical Pharmacology, Seth GS Medical College and KEM Hospital, Department of Anesthesiology, Tata Memorial Centre, Homi Bhabha National Institute, Mumbai, Maharashtra, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India

Abstract

Randomized controlled trials are the gold standard for determining the efficacy of a new intervention. Trials conducted for regulatory approval of an intervention compare the effect of the intervention with the standard of care or placebo to demonstrate efficacy. Randomization attempts to ensure that all known and unknown confounding factors are evenly distributed between the groups, and that the groups will be comparable at the end of the study, so that any inter-group differences in outcomes can be attributed to the intervention. However, in reality, intercurrent events may impact the assessment and subsequent interpretation of the outcome of interest. To address this, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) in 2017, released an addendum to the E9 guideline (ICH E9 R1) putting forth the concept of Estimands and Sensitivity Analysis in Clinical Trials. This addendum addresses how these intercurrent events are to be handled using the Estimand concept, which is now expected to be detailed in a separate section of the study protocol. In this paper, we discuss what estimands are, and their likely impact on how regulatory trial protocols and their statistical analyses plans are written and implemented. We also look at the application of the concept of estimands to routine clinical practice.

Keywords: Clinical trials, drug development, estimand, hypothetical, regulatory

BACKGROUND

The International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) has guidelines for various aspects of medications, namely efficacy (E), safety (S), quality (Q), as also multidisciplinary guidelines. Of these, the ICH E9 guideline pertains to Statistical Principles for Clinical Trials. In August 2017, ICH released an addendum to E9 (ICH E9 R1) which put forth the concept of “Estimands and Sensitivity Analysis in Clinical Trials.” The concept soon found acceptance with the regulatory agencies globally. The US FDA, for example, stated that this addendum document, when finalized, would reflect the agency’s current thinking. The final version of the addendum was adopted on November 20, 2019. Since then, there has been a lot of activity in this area including, for example, the setting up of working groups (e.g. Oncology Estimand Working Group). In July 2020, Health Canada endorsed the principles and suggested adoption of the practices laid down in this addendum. Thus, this concept is expected to soon find application worldwide.
The present article discusses what estimands are, how their use is likely to change the way in which protocols for regulatory clinical trials are written and statistical analyses are carried out, as also the relationship of estimands with intention-to-treat (ITT) analyses. It also briefly touches upon how clinicians can apply the concept of estimands in assessing how the results of a particular trial relate to their clinical practice. However, this article deals with only the major principles and is not intended to be a comprehensive treatise on the subject.

**WHY ESTIMANDS ARE NEEDED, THEIR PREMISE AND A SIMPLE DEFINITION**

Randomized controlled trials form the cornerstone of regulatory approval of drugs, and evidence-based medicine and policy. Regulatory trials in drug development set out to estimate the magnitude of effect of an intervention (drug/device/surgical procedure) relative to a placebo/standard of care. These studies are designed to produce treatment groups that are balanced to the extent possible and it is expected that all the participants will complete the trial as per the protocol and will provide complete data. The ultimate objective is to obtain an accurate estimate of the measure of effect of the intervention being studied. However, in real life, more often than not, intercurrent events impact the measurement and consequently the interpretation of the outcome of interest. Some examples of intercurrent events are described in Table 1. Occurrence of intercurrent events leads to each patient in a clinical trial following a slightly different path. Each such alternative path leaves its own impact on the measurement of the true benefit of the intervention being evaluated.9

Intercurrent events can be related to the disease or to the specific intervention being studied, or may be unrelated (for example a participant relocating to another city). Potential intercurrent events that may occur during a trial can thus vary with the nature of disease or population studied, and with the planned intervention and the alternative or additional treatment options available to the study participants. The estimand framework requires that the sponsor and/or investigators think through about the various possible intercurrent events and their combinations well in advance and address these a priori while planning the study and explicitly state these in the protocol in a separate section, called the estimands section. Thus, an estimand is a way for the clinical study protocol to address how intercurrent events will be dealt with. To paraphrase ICH E9 R1, an estimand is “a precise description of the treatment effect reflecting the clinical question posed by the trial objective that summarizes outcomes in the same set of patients under diverse treatment conditions.” The idea is to add precision to the research question so that the quality and conduct of the study are improved and ultimately the design, conduct, data analysis, and interpretation are aligned with the study objectives, despite the occurrence of intercurrent events that preclude a perfect match between the conduct of the study and the study protocol. Two related terms are estimate and estimator. The former refers to a numerical value that we derive at the end of the study, whereas the latter refers to the analytic method leading to the numerical value.

A particular post-randomization event may or may not be an intercurrent event depending upon the therapeutic area and research question. Death, for example, would be an intercurrent event in a study evaluating efficacy of an analgesic in relieving pain, whereas it would be an endpoint in an oncology study. It is important to note that a single study can have more than one estimand, each providing a slightly different viewpoint and conclusion.

**THE FOUR ELEMENTS OF THE ESTIMANDS FRAMEWORK**

An estimand has four inter-related attributes: the population, the variable, the specification of how to account for intercurrent events, and the population level summary [Table 2]. We discuss these using two examples as outlined below.

**Example 1**

Boonen et al[7] evaluated the effect of zolendronic acid on fracture risk in men with osteoporosis. This was a double-blind,
Gogtay, et al.: Understanding estimands

Table 2: Understanding the estimand framework in a diabetes trial

| Element                  | Patient or outcome characteristic                                                                 |
|--------------------------|---------------------------------------------------------------------------------------------------|
| Population               | Adult patients with type 2 diabetes                                                                |
| Variable                 | Change from baseline in HbA1c at week 26                                                            |
| Intercurrent events      | Patients who took rescue medication or prematurely discontinued treatment*                          |
| Population level summary | Between group difference of the change from baseline in HbA1c at week 26                            |

*See Table 1 for other examples of intercurrent events. HbA1c = Glycated haemoglobin

placebo-controlled trial where men (50–85 years) with primary or hypogonadism-associated osteoporosis were randomly assigned to receive intravenous infusions of either 5 mg of zolendronic acid or of placebo at baseline and at 12 months. The primary outcome measure was the proportion of participants with one or more new morphometric vertebral fractures over a period of 24 months.

Element 1: Population
The target population for the research question is the population of people who would satisfy the eligibility criteria. This population refers to the wider pool of people who would be eligible for this trial rather than those who were actually recruited into the study. In this example, it is the men over the age of 50 years and under 85 years with osteoporosis that is either primary in origin or is the result of hypogonadism.

Element 2: Variable/endpoint
This is the study endpoint (or primary or secondary outcome) that is clearly defined in the protocol. In this case, it is the proportion of participants who had developed one or more new morphometric vertebral fractures over a period of 24 months, as assessed using radiographs taken at 12 and 24 months.

Element 3: Intercurrent events
An intercurrent event is defined as an event that occurs after the study intervention has been initiated and which either precludes the observation of the outcome variable, or affects its measurement or interpretation. In this study, several events of importance occurred in both groups, which precluded perfect adherence to the study protocol. These were: patients who discontinued the intervention (did not take the second dose), withdrew consent, died prematurely, had an adverse event (and discontinued both treatment and follow-up, or did not take the second dose but had radiographs at the intended time-points) or were lost to follow-up (took both doses but did not return for 24-month radiograph). Occurrence of some of these events (e.g., premature death, loss to follow-up) precluded the measurement, and that of others (e.g., treatment discontinuation) could affect the value of variable/endpoint of interest, i.e., the occurrence of new morphometric fractures.

Element 4: Population level summary
This is the variable based on which the difference between the two interventions will be decided. In this example, this would be the between group difference, i.e., the difference in the proportion of men with one or more new morphometric vertebral fractures over a period of 24 months between the two groups (treated with bisphosphonates and untreated).

Conventionally, all these components used to be written in the section on statistical analysis. Thus, estimands have been called as “old wine in new barrels.” However, per the ICH addendum, these elements are now expected to be described in the protocol in a separate section, called the estimand section. The set of intercurrent events described will, as may be expected, depend on the therapeutic area being studied and the objective of the study. Let us use a few examples to understand what the estimands section looks like in a study protocol.

Example 2: The PIONEER Studies from the PIONEER program
The PIONEER drug development program (which has multiple studies) to evaluate the role of semaglutide (first-in-class GLP-1 agonist) in the treatment of diabetes mellitus. These studies being among the initial studies that have incorporated the concept of estimands in their planning, conduct, analysis and interpretation, have been frequently used in literature as examples to explain the conceptual framework of estimands. The estimand framework for the PIONEER studies is given in Table 2.

ESTIMAND STRATEGIES FOR ADDRESSING INTERCURRENT EVENTS
When intercurrent events are expected to occur in a trial, one of the several strategies, as discussed below, can be used to address these [Table 3]. If several different intercurrent events are considered possible, different strategies can be used to address each.

Treatment policy strategy
In this strategy, the occurrence of intercurrent event is ignored and measurements of the variable of interest are used as such. For example, in the PIONEER 1 study, where glycated hemoglobin (HbA1c) was the endpoint of interest, the data on HbA1c could be analyzed regardless of whether a particular patient took (or did not take) any rescue medication.
In the protocol for this study, the primary estimand was defined as “treatment difference (oral semaglutide vs. placebo) at week 26 for all randomized subjects regardless of adherence to the allocated treatment and initiation of rescue medication.”[9] Thus, this estimand considered two post-randomization events (lack of adherence and use of rescue medication) as irrelevant to the final analysis. In other words, the analysis did not make any statistical adjustment for the use of rescue medication or discontinuation of the study drug.[9] The use of this strategy as the “primary” estimand in this trial was predicated on the fact that the use of rescue medication reflects the usual clinical practice and the adherence to treatment (or lack thereof) reflects the real-life behavior of the target population.

Let us take another (hypothetical) example of a study by Mitrinović et al.[8] that evaluated the analgesic effect of a medication, where self-administration of rescue medication was an intercurrent event and reduction in pain as measured on a visual analogue score (VAS) was the outcome of interest. In this strategy, the VAS at the end of the study could be used as is, ignoring the intercurrent event of some patients in each group having taken the rescue pain medication.

This estimand in both these examples would correspond to the “ITTT” principle where all randomized patients are analyzed irrespective of some intercurrent events, namely lack of adherence to the study treatment or taking additional medications.

However, this estimand would be inappropriate if a large proportion of study participants leave the study and are not available for the final measurement which defines the endpoint.

**Composite strategy**

In this strategy, an intercurrent event is integrated into one or more clinical outcome measures. In other words, the intercurrent event becomes a part of the endpoint definition itself. This can be understood with two examples.

The SYNAPSE study for nasal polyps (ClinicalTrials. Gov identifier: NCT03085797) is aimed at evaluating mepolizumab vs. placebo on the nasal polyp score and the nasal obstruction VAS (co-primary endpoints) over a 48-week period.[13] An intercurrent event of interest in this study is surgery for nasal polyps (in event of lack of clinical improvement), and it is estimated that 40% of participants in the placebo group will undergo surgery. The composite estimand in this study is stated as “Improvement in nasal polyps score of ≥1 point and completion of treatment period without surgery.”

Similarly, in PIONEER 1,[9] rescue medication is the intercurrent event of interest. Thus, an endpoint such as

| Estimand strategy | Measurement of interest |
|-------------------|-------------------------|
| Treatment policy strategy – value of the outcome of interest regardless of the occurrence of the intercurrent event - the intention to treat analysis | Proportion of patients who achieved a goal of HbA1c <7% regardless of whether they took (or did not take) rescue medication[9] Therapeutic area: diabetes |
| Composite strategy – value of the variable with the intercurrent event being woven into the outcome variable | Proportion of patients with improvement in nasal polyps score of ≥1 point and completion of treatment period without surgery[13] Therapeutic area: nasal polyps |
| Hypothetical strategy – assumes that the intercurrent event did not happen in patients who were randomized | Proportion of patients who would achieve a goal of HbA1c <7% had they not taken rescue medication in a diabetes trial. Since patients in these studies are likely to take rescue medication, this value of <7% (in those who did take rescue medication) would be calculated using a specific statistical technique (modelling method) that is clearly stated in the estimand section of the protocol Therapeutic area: diabetes |
| Principal stratum strategy – measurement of the variable of interest is a subgroup of patients not likely to need rescue medication or not likely to discontinue treatment | Proportion of patients in a diabetes study (that includes both prediabetics and diabetics) who achieve a target HbA1c of <5.7%. If we look at the prediabetic patients (HbA1c between 5.7% and 6.4%) as a separate sub group, this group would be the one not likely to have needed rescue medication Therapeutic area: diabetes – this is a hypothetical example. Note that prediabetics are extremely unlikely to need rescue medication in any diabetes study |
| While on treatment strategy – measurements up until the time of the event particularly when the measurements are repeated at multiple time points | If patients of NASH are treated with a new intervention, serial liver biopsies are carried out at six monthly intervals to assess its effect. Fibrosis regression of at least two stages without worsening of NASH is the outcome of interest. If some of these patients are well controlled diabetics and their diabetes worsens while on treatment leading to medication discontinuation (with the new intervention), the results of the liver biopsies up until the discontinuation can be considered Therapeutic area: GI medicine – this is a hypothetical example |

HbA1c = Glycated haemoglobin, GI = Gastrointestinal
“the proportion of patients in both groups who reached an absolute HbA1c value of <7% at the end of the study without the use of rescue medication and who continued to use the intervention of interest throughout the duration of the study” would constitute the composite strategy.

In the two examples described, using the “composite” strategy, the intercurrent events of interest (surgery and rescue medication, respectively) have been neatly woven into the endpoints (of nasal polyps score and HbA1c reduction, respectively).

**Hypothetical strategy (also called as Trial Product Estimand)**

This envisages a scenario wherein it is presumed that an intercurrent event (for example, the use of rescue medication or discontinuation of follow-up for an unrelated reason in a diabetes study) would not have occurred. Thus, assumes that a particular type of intercurrent/post-randomization event did not happen in the patients who consented to participate.

An example can be found in the secondary estimated in the PIONEER 1 study, which evaluated “treatment difference (oral semaglutide vs. placebo) at week 26 for all randomized subjects if all subjects adhered to treatment and did not initiate rescue medication.” In the patients who prematurely started a rescue medication or discontinued follow-up, the analysis used partial data available for the particular participant (i.e., measurements up to the intercurrent event) and a mixed model for repeated measurements to estimate what the likely value of HbA1c would have been at 26 weeks, if the intercurrent event had not occurred and he had followed the pattern of other patients who continued follow-up. This estimand would thus avoid the confounding effect of use of a rescue medication or of discontinuation of follow-up on the endpoint.

However, one needs to decide which intercurrent events are suitable for the use of this strategy. For an inter-current event such as lack of follow-up data because of an unrelated reason (e.g., moving away from the study area), the use of modeled value may be reasonable. However, if the intercurrent event is a change in treatment due to a lack of response, this strategy may not be the most appropriate.

**Principal stratum strategy**

This strategy considers measurements in a subgroup or target population where the intercurrent event(s) is not likely or less likely to occur. This can be done by using an appropriate study design.

An example of this is found in the protocol for the currently-ongoing SONAR study, which is a double-blind, randomized, placebo-controlled trial on the effect of atrasentan, a selective endothelin A receptor antagonist, on the occurrence of renal events in patients with type 2 diabetes and chronic kidney disease.[13] This study initially enrolled 4711 patients with type 2 diabetes who had an estimated glomerular filtration rate within 25 to 75 mL/min/1.73 m² and urinary albumin-to-creatinine ratio between 300 and 5000 mg/g. In the first phase, all the participants received atrasentan, in a dose of 0.75 mg/day for 6 weeks. Based on the response to this, the participants were then divided into two sub-groups – responders and non-responders. Only the responders (n = 2648), defined as those with ≥30% reduction in urinary albumin-to-creatinine ratio, weight gain of less than 3 kg, an increase of no more than 0.5 mg/dL and of 20% from baseline in serum creatinine, and with absolute brain natriuretic peptide <300 pg/mL at the end of the 6-week enrichment period, were considered for inclusion in the eventual study.[13] The selection at the end of the initial phase helped exclude patients who would have been at risk of developing certain intercurrent events, such as marked weight gain, rise in brain natriuretic peptide or rise in serum creatinine, which would have led to discontinuation of the drug of interest – and hence obviate the interference that these events could cause with the interpretation of the effect of the drug.

As a variation on this theme, it may be possible to treat different types of participants (e.g., the responders and non-responders in the first phase in the study above) as two distinct strata, with patients in each being separately randomized to an intervention versus placebo. Such a design would allow one to assess the effect of the intervention of interest separately in different types of participants.

**While-on-treatment strategy**

The idea in this strategy is to evaluate treatment effect prior to the occurrence of the intercurrent event. This approach is particularly useful when the endpoint is measured multiple times during the course of the study. This strategy uses measurements up until the time of the event, and the trial is designed such that measurements are stopped post the intercurrent event.

For example, in a study on Plasmodium falciparum malaria, parasite clearance time and fever clearance time are measured daily on days 1–3 and then on days 7, 14, 28, and 42.[14] One intercurrent event of interest is the clinical failure (due to the problem of drug resistance) which would
lead to the administration of rescue therapy (quinine). Analysis of measurements of parasite and fever clearance up until the use of rescue medication would constitute the “while-on-treatment” estimand strategy. In this strategy, measurement of endpoint is discontinued once the intercurrent event occurs.

Similarly, let us think of a study in which patients with nonalcoholic steatohepatitis treated with a new intervention, and where serial liver biopsies are carried out at 6-month intervals to assess the effect of the intervention. If the new intervention leads to worsening of diabetes control necessitating change of treatment, the data for liver biopsies done up to the time when the treatment was changed can be analyzed and further liver biopsies can be discontinued.

SENSITIVITY ANALYSIS

To quote ICH E9 R1, sensitivity analysis is defined as “a series of analyses conducted with the intent to explore the robustness of inferences from the main estimator to deviations from its underlying modeling assumptions and limitations in the data.” We may recall that the estimator is the analytical method that yields the estimate or the final numerical value. All estimands discussed thus far have statistical assumptions on which they are based. One could try to vary these statistical assumptions to see what effect this has on the estimate of the drug effect. For instance, in the example used above for the “hypothetical strategy,” one could use different mixed effects models to see whether the estimate of drug effect is sensitive to the choice of model. If different models lead to similar results, the estimate is more likely to be reliable.

HOW WILL CLINICIANS UNDERSTAND AND USE ESTIMANDS IN THEIR PRACTICE?

Estimands based on different strategies may have application in different clinical settings or may answer somewhat different clinical questions.

Estimands based on the treatment policy strategy and the composite strategy, which account for the addition of rescue medication, discontinuation/switch of medication due to adverse events, etc., will give a real-world understanding of the use of a new product.

The principal strategy estimand can help inform the clinician about the effect of a drug in the subset of patients who are likely to continue taking the drug and hence are the most likely to benefit from it.

The while-on-treatment strategy will help clinicians assess the impact of an intervention up to a certain point in time and allow for an understanding of how the intervention has (or has not) helped the patient up until that point.

The hypothetical strategy perhaps has the least importance or relevance to clinicians as it does not depict the real-world scenario.

SUMMARY AND CONCLUSIONS

The importance of the ICH E9 R1 addendum lies not in its novelty but rather in the clarity that it provides in assessing trial results, through the use of guidance and a framework for addressing intercurrent events. The choice of estimand(s) in a trial will be driven by the therapeutic area, characteristics of the intervention, alternative therapies available and the characteristics of the population being studied. While not intended to be a gold standard yet for clinical trials, the estimands framework helps think about various intercurrent events at the planning and design stage itself, so that these do not impact conclusions about whether or not an intervention actually works or is likely to work in the real-world setting (efficacy vs. effectiveness).

Finally, it is important to remember that use of more than one estimand for a given study can add value to the latter. Not too often, even meticulously-conducted regulatory clinical trials are criticized soon after their publication, throwing the drug into disrepute and reducing the eventual use of an effective intervention. The estimands framework allows the investigators to think a priori of some of the likely issues, which would otherwise later be considered as shortcomings, and plan the trial estimands better. Furthermore, various stakeholders (physician, patient, regulator, pharmaceutical industry, payer) have different expectations from a clinical trial, and hence one can plan for multiple estimands – each serving a different stakeholder group. This can improve the understanding of the effect of a particular intervention across these diverse stakeholder groups – helping them look at the study results from their specific viewpoints and with greater respect, rather than through the prism of limitations conferred by inevitable intercurrent events.

Given their several advantages, the use of estimands should see a rapid increase over the next few years, and all of us need to be ready to adapt to these.

Financial support and sponsorship
Nil.
Conflicts of interest
There are no conflicts of interest.

REFERENCES

1. Available from: https://www.ich.org/page/efficacy-guidelines. [Last accessed on 2020 Dec 08].
2. Available from: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf. [Last accessed on 2020 Sep 25].
3. Available from: https://www.fda.gov/media/108698/download. [Last accessed on 2020 Oct 06].
4. Available from: https://oncoestimand.github.io/oncowg_webpage/oncoestimand.html. [Last accessed on 2020 Oct 06].
5. Available from: https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/applications-submissions/guidance-documents/international-conference-harmonisation/efficacy/notice-ich-e9-r1-defining-appropriate-estimand-clinical-trial-sensitivity-analysis.html. [Last accessed on 2020 Oct 06].
6. Degtyarev E, Zhang Y, Sen K, Lebwohl D, Akacha M, Hampson LV. Estimands and the patient journey: Addressing the right question in oncology clinical trials. JCO Precision Oncology 2019. doi.org/10.1200/PO.18.00381.
7. Boonen S, Reginster JY, Kaufman JM, Lippuner K, Zanchetta J, Langdahl B, et al. Fracture risk and zoledronic acid therapy in men with osteoporosis. N Engl J Med 2012;367:1714-23.
8. Mitroiu M, Oude Rengerink K, Teerenstra S, Péauté F, Roes KC. A narrative review of estimands in drug development and regulatory evaluation: old wine in new barrels? Trials 2020;21:671.
9. Available from: https://clinicaltrials.gov/ProvidedDocs/30/NCT02906930/SAP_001.pdf. [Last accessed on 2020 Nov 09].
10. Aroda VR, Saugstrup T, Buse JB, Donsmark M, Zacho J, Davies MJ, et al. Incorporating and interpreting regulatory guidance on estimands in diabetes clinical trials: The PIONEER 1 randomized clinical trial as an example. Diabetes Obes Metab 2019;21:2203-10.
11. Keene ON. Strategies for composite estimands in confirmatory clinical trials: Examples from trials in nasal polyps and steroid reduction. Pharm Stat 2019;18:78-84.
12. Heerspink HJL, Andress DL, Bakris G, Brennan JJ, Correa-Rotter R, Dey J, et al. Rationale and protocol of the study of diabetic nephropathy with aRasentan (SONAR) trial: A clinical trial design novel to diabetic nephropathy. Diabetes Obes Metab 2018;20:1369-76.
13. Heerspink HJL, Parving HH, Andress DL, Bakris G, Correa-Rotter R, Hou FF, et al. Atrasentan and renal events in patients with type 2 diabetes and chronic kidney disease (SONAR): A double-blind, randomised, placebo-controlled trial. Lancet 2019;393:1937-47.
14. Okafor HU, Shu EN, Oguonu T. Therapeutic efficacy and effect on gametocyte carriage of an artemisinin and a non-based combination treatment in children with uncomplicated P. Falciparum malaria, living in an area with high-level chloroquine resistance. J Trop Pediatr 2010;56:398-406.