Estimands – closing the gap between study design and analysis

Helen Bridge[1] and **Thomas M. Schindler**[2]
1   AstraZeneca, Cambridge, UK
2   Boehringer Ingelheim Pharma, Biberach, Germany

Correspondence to:

Helen Bridge
AstraZeneca
Cambridge
UK
Helen.Bridge@astrazeneca.com

Abstract

Estimands represent a new way to look at key aspects of clinical research and will become increasingly important for medical writers. Estimands are detailed definitions of quantities to be estimated using clinical trial data, which make allowance for events that happen after randomisation. Such post-randomisation events include, for example, treatment discontinuation due to poor tolerability or lack of efficacy, and use of rescue medication. Through a worked example, this article elucidates several different kinds of estimands and shows how the estimands approach has the potential to improve the quality of clinical research. Estimands foster a more complete alignment of study objectives, study design, study conduct, data analysis, and interpretation of results.

The word *estimand* may look like a spelling mistake, but it actually represents a new paradigm in clinical research. With the new term comes a set of concepts that will change the way we perform clinical studies, particularly pivotal phase III studies. Estimands are not really a statistical idea, but rather one that pertains more generally to the evaluation of clinical trial results.

Clearly the word estimand is related to "estimate": an estimand is a clinical entity or parameter that is estimated by performing a

clinical study. In other words, an estimand is the *target* of estimation; the aim is to capture this target of estimation as precisely as possible. The concept of estimands is the subject of lively discussion in the statistical community and is outlined in a recent draft addendum to the International Council for Harmonisation (ICH) guideline on statistical principles for clinical trials, ICH E9.[1] In the months after its release for public consultation at the end of August 2017, a number of organisations and individuals submitted comments and suggestions on the draft addendum to the EMA.[2] The ICH E9(R1) Expert Working Group recently released a collection of training materials that elaborate on the content of the addendum and make suggestions on its implementation.[3] The final version of the addendum is expected in 2019.

Randomisation and intercurrent events

Given the effort and cost involved in conducting a clinical study, we want to be sure it produces objective results that have not come about by any systematic error that shifts the results in a certain direction. A central method to avoid bias is randomisation. By randomly assigning the patients in a study to two parallel treatment groups, we ensure that the two groups are comparable at study start with respect to both known (measured) and unknown characteristics.

Then we can safely ascribe any effect we see to the treatment we are investigating – or at least, that is the common belief. In fact, however, this is only true if the initial randomisation is maintained during the study – and that is often not the case because of "intercurrent events",[4] for which a better term would be "post-randomisation events" (the two terms are used interchangeably in this article).[2] These are any events that happen to patients during a study and that may affect the results. In particular, the following intercurrent events are important: patients die, they stop taking the study medication because they experience side effects or because they feel they are having no benefit from the treatment, or they take additional medication that will interfere with the efficacy endpoints (Figure 1).

Randomisation ensures that the variation among individuals is similar in the two treatment groups at baseline. However, each individual patient is likely to experience different inter-current events depending on which treatment he or she receives. This may result in differences in the rates and timing of intercurrent events between the treatment groups. If we exclude all patients who experience intercurrent events from the analysis then we may, at the time when the study results are determined, no longer have treatment groups that are comparable. This is why, until now, industry guidance (ICH E9) has recommended performing an intention-to-treat
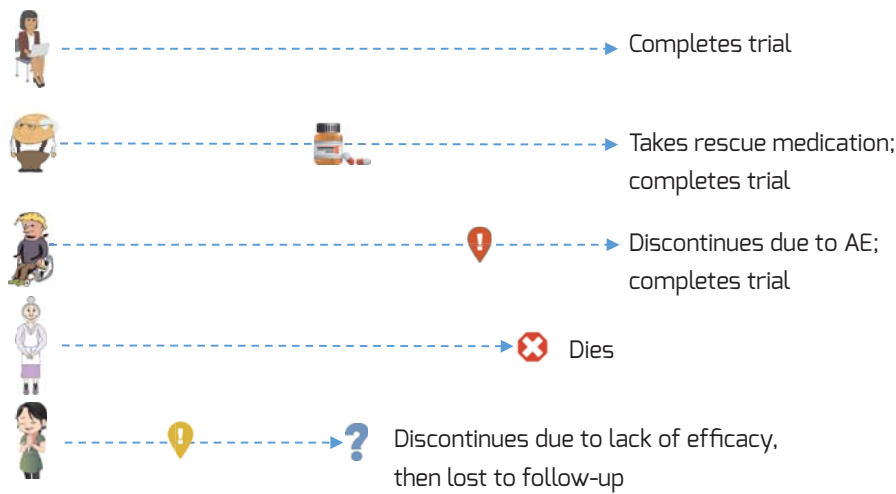
Figure 1. Examples of post-randomisation events that may occur in a group of patients

(ITT) analysis on all randomised patients, or at least, as close to all randomised patients as possible. The new addendum to ICH E9 recognises that this guiding principle has its limitations.[4]

Effects of post-randomisation events

The potential effect of post-randomisation events is best illustrated with an example. Assume we have a study in patients with type 2 diabetes and we want to compare two treatment groups: one group receives wonderdrug (WD) and the other group receives placebo, both in addition to background therapy. We want to measure the treatment effect by comparing the reduction in glycated haemoglobin (HbA1c), a long-term marker of blood glucose levels, from study start to Week 26.

In trials in type 2 diabetes it is standard to make rescue medication available to patients whose blood glucose level is not adequately controlled with the study treatment. This means that patients whose blood glucose exceeds a predefined limit are allowed to take additional antidiabetic medications alongside the study treatment. This is done because high blood glucose increases the risk of complications such as cardiovascular problems or damage to the nerves, kidneys, or eyes. It would not be ethical to require patients to continue in the trial with excessive blood glucose levels.

However, from a scientific point of view, the use of rescue medication in a trial complicates the evaluation of the treatment effect. The question

is what to do with the data when patients start taking rescue medication. Do we continue to take efficacy measurements in these patients, and do we include such measurements when we calculate the treatment effect?

Clearly our decision with regard to trial design will have consequences for how we need to interpret the results. Up to now, such consequences have not always been considered at the trial design stage.[5] For example, if we plan the trial in such a way that data are not collected from patients after they have started rescue medication (e.g. because such patients are withdrawn from the trial), then we may end up with only a small number of patients with data at Week 26. Our options at the analysis stage will then be limited; a full ITT analysis will not be possible. If, on the other hand, we collect and use data from all patients, even after rescue medication use (i.e., the ITT approach), then the measured values will reflect both the effect of the study treatment and the effect of the rescue medication, resulting in a comparison of WD plus rescue medication and placebo plus rescue medication. Depending on which option for the collection and analysis of data is chosen, the precise definition of the treatment effect (or estimand) will differ. Although using the ITT approach helps to ensure statistical validity, the estimate of treatment effect it produces may not be clinically meaningful because the effect of WD will be "blurred" by the effect of rescue medication. This situation has been described in terms of a trade-off between "having a precise answer to a less relevant question or an approximate answer to the most

relevant question".[6] The estimands discussion makes clear that this trade-off can be made in a variety of different ways.[3]

The new approach: Estimands

Rather than arriving at a particular estimand implicitly and haphazardly as a consequence of choices about data collection and statistical analyses, the ICH E9 addendum suggests that we should consider explicitly and up front the various scientific questions that the trial data could be used to address. Using estimands allows us to see intercurrent events as a source of important additional information on the efficacy and safety of an investigational treatment, rather than treating them as a nuisance or complication.[7]

We can then choose which questions – and hence estimands – are the most meaningful in our clinical context and which are most relevant for patients, their doctors, regulators, and payers. The disease setting and aim of treatment will affect the choice of estimands.[3] In many settings, a single estimand is unlikely to meet the different needs of all stakeholders.[8, 9] It has even been proposed that the most helpful way to provide physicians and patients with the information they need about a treatment would be to include a (lay) description of several estimands in the prescribing information.[5, 6]

Compared with endpoints as currently defined in clinical trial protocols, estimands are more detailed definitions of the quantity to be estimated and comprise four interrelated attributes, described in the ICH E9 draft addendum as follows:

- **Population**: Which patients are targeted by the scientific question?
- **Variable/endpoint**: Which quantity needs to be obtained for each patient to address the scientific question?
- **Intercurrent/post-randomisation events**: How are these to be accounted for to reflect the scientific question?
- **Population-level summary**: Which summary statistic (e.g. mean or median) for the variable will be the basis for comparing the treatments?

In our diabetes example many different estimands are possible, and the situation would become even more complicated if we were to consider other kinds of intercurrent events (e.g., deaths and discontinuations due to adverse events) in addition to rescue medication use.[3,10] Estimands could also be defined for answering questions

Table 1. Three possible estimands for a trial in type II diabetes

	Endpoint Variable	Intercurrent/Post-randomisation Event
Estimand 1 (treatment policy)	Change in HbA1c from baseline to Week 26	Consider all data regardless of rescue medication use
Estimand 2 (on-treatment)	Change in HbA1c from baseline to Week 26 or to the last value before initiation of rescue medication	Data after initiation of rescue are not considered
Estimand 3 (hypothetical)	Change in HbA1c from baseline to Week 26	Data after initiation of rescue are modelled as if no patients took rescue medication before Week 26

The estimand descriptors given in brackets are the terms used in the draft ICH E9 addendum.

related to safety, e.g. How long are patients able to remain on the treatment before discontinuing due to adverse events?

To keep things simple, we will look at just three types of estimands for the treatment effect that deal with rescue medication use in different ways (Table 1). (The draft version of the ICH E9 addendum actually defines five different types of estimands.)[1] All three estimands define the same patient population (i.e., the one in which it is planned to use WD after approval, as reflected by the trial inclusion and exclusion criteria), and all use the difference in mean change in HbA1c values between the treatment groups as the "population-level summary". The differences among the estimands lie in the precise definition of the variable to be used as the primary endpoint and in the handling of the intercurrent event "use of rescue medication".

Estimand 1 is the estimand corresponding to the ITT analysis described in the previous section, and requires that we use the HbA1c data from all patients at Week 26, including those who have started rescue medication (Figure 2). If WD is effective at lowering HbA1c, then we can expect that the use of rescue medication will be more frequent in the placebo group. The treatment effect we estimate at Week 26 will then be the difference between the effect achieved by WD, occasionally with additional rescue medication, and the average effect seen in a placebo group where many patients are taking rescue medication known to be effective in reducing HbA1c.

Given the blurring of the efficacy of WD by the use of rescue medication in the control arm, we may end up with a modest difference between

the treatment groups that underestimates the true difference between WD and placebo. On the other hand, we will obtain a result that reflects clinical practice "out there", because it is very likely that some patients in clinical practice will require additional medication, whether they are taking WD or other standard antidiabetic medications. Such an approach is called a "treatment policy estimand", and this analysis is likely to be of particular importance to payers and reimbursement agencies who want to know the effectiveness of WD in the real world. This is also sometimes called an "effectiveness estimand".[11]

Analysing all patients according to the treatment they were randomised to, rather than

the treatment they actually received, helps to ensure that the treatment effect is not overestimated and that statistical tests produce valid results. In 2011, a US FDA reviewer used precisely this argument to suggest that the most valid way to analyse data for the new antidiabetic drug dapagliflozin was to use all data, including values from patients taking rescue medication, in the statistical model.[12] This incident was a trigger for the estimands debate.[4,10]

Estimands 2 and 3 attempt in different ways to capture the effect of WD itself without blurring it by the use of rescue medication.

Estimand 2 considers all data up to Week 26 or the time when rescue medication was initiated (Figure 2). It estimates the effect of the treatments until rescue was needed or until Week 26 for patients who did not need rescue medication. If WD works, few patients in this group will need rescue medication, and those who do need it are likely to need it late in the trial. Conversely, in the placebo group many patients will need rescue because their background medication will not control blood sugar effectively and they are likely to need to initiate rescue medication soon after study start. Using the last recorded HbA1c value before start of rescue medication means this analysis will use values for many patients, particularly in the placebo group, at a time point when HbA1c values are likely to be high. This estimand will therefore have a tendency to overestimate the

Estimand 1 = change from BL to W26 regardless of rescue medication use:

Estimand 2 = change from BL to last value before rescue medication:
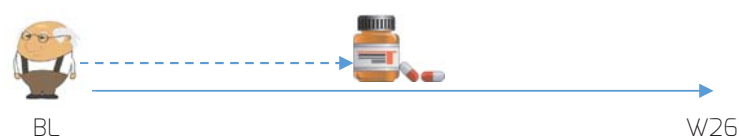
Estimand 3 = change from BL to W26 as if no patients took rescue medication:

Figure 2. How three estimands account for rescue medication use. BL: baseline; W26: Week 26

How will estimands affect medical writers' work?

- Medical writers may come across estimands while writing study protocols. Estimands will need to be described for the primary and key secondary endpoints. Study objectives, study endpoints, and the analysis methods for the results will need to be closely aligned and described in detail. To this end, estimands will need to be agreed upon cross-functionally at the early stages of protocol development. Medical writers will need to understand estimands to facilitate this process.
- In the study protocol and the case report form, the reasons for discontinuation of treatment will need to be captured with greater granularity. In the informed consent form, patients will need to be asked for consent for data collection to continue if they decide to discontinue treatment.
- In clinical study reports, the methods sections for the description of study design and objectives, choice of endpoints, and analysis strategy will need to outline the estimands chosen. The results sections will need to be organised around estimands in addition to endpoints, and will need to describe the occurrence and timing of post-randomisation events. Reports will also need to include discussion of any limitations of the chosen estimands and of how unforeseen post-randomisation events were handled in the analysis.
- Writers of clinical submission documents will need to describe estimands comprehensively to justify the choice of patient population and endpoints for the proposed drug label.

Guidance on the documentation of estimands is included in the ICH E9 addendum training materials.[3]

effect of WD.[10] Such an approach is categorised as an "on-treatment" estimand because it estimates the response to treatment prior to the occurrence of the intercurrent event.[1]

In effect, Estimand 2 corresponds to a last-observation-carried-forward (LOCF) analysis. Although LOCF has been widely used to deal with rescue medication in type 2 diabetes trials, it is a problematic approach because it results in estimates that are biased.[13] Bias may also be a disadvantage of Estimand 2, and the resulting estimate will be difficult to interpret because it is based on a comparison of data at widely different time points in the two treatment groups. However, in some clinical trial settings, such as the evaluation of palliative treatments in end-of-life care, where death is an expected post-randomisation event, a while-on-treatment estimand may be the most clinically meaningful one.[3]

Estimand 3 provides a very different and less intuitive, yet interesting approach. With this estimand we estimate the treatment effect that would be seen if no patients took rescue medication (Figure 2). The analysis for Estimand 3 will include only values from patients who have not (yet) started rescue medication; HbA1c values will be counted as missing from the point when a patient starts rescue medication. An appropriate method for handling missing data through statistical modelling (e.g. multiple imputation [MI] or mixed models for repeated measures [MMRM]) will need to be used.[10,13] The resulting estimate will reflect both what actually happened in patients who reached Week 26 without rescue medication and what the data collected before rescue medication suggest might

have happened by Week 26 in the remaining patients if they had continued without rescue medication. This estimand is hypothetical at the level of a group of patients: it relies heavily on the modelling of the data for a large proportion of patients and will therefore never fully reflect a "real-life" situation. The usefulness of this type of estimand has been contested, particularly by health technology assessment agencies responsible for assessing the value of a treatment in actual clinical practice.[2] The ICH E9(R1) Expert Working Group advises that hypothetical estimands should be based on clinically reasonable situations that are clearly specified in the clinical study protocol.[3] Based on appropriate statistical modelling, Estimand 3 is likely to provide a less biased answer than Estimand 2 to the question that is crucial to individual patients: "If I take this drug as part of my treatment regimen, without adding any further drugs, what effect can I expect to see after 26 weeks?"

Estimands and trial design

We have tried to make it clear that estimands will help researchers to formulate more clearly what they really want to get out of a clinical study. The traditional approach does not adequately take into account the effects of intercurrent events on the primary endpoint measure.[4] As demonstrated in the example of rescue medication use in a type 2 diabetes trial, depending on how we account for such events, we may be aiming to estimate the effect of the study drug itself or we may actually be evaluating a treatment policy.

In the past it was often the case that clinical researchers tried to elucidate what exactly they

had evaluated after a study had been completed. This is surely not the ideal situation because very little can be done after the fact. For example, once the decision has been taken not to collect data after initiation of rescue medication, this cannot be reversed after trial completion.

The paradigm shift introduced by the idea of estimands involves a different sequence of activities (Figure 3).[3,5,9] Clinical researchers first need to think about the objectives of the trial (i.e., what the trial is meant to show). An objective could be to demonstrate the effectiveness of a drug in reducing HbA1c in patients with type 2 diabetes. Researchers then need to consider the precise scientific questions of interest to be addressed and to choose estimands that answer these questions. In order to ensure that key stakeholders' needs are met, this choice should be made in discussion with regulatory authorities and in accordance with available guidance.[2,3] Indication-specific guidelines on the appropriate use of estimands are likely to become available in the future. Once the estimands have been defined, the trial can be designed in such a way that all the necessary data are collected, and the statistical analysis methods can be chosen to address the estimands of interest. For many estimands relevant to patients and physicians, it will be necessary to record reasons for treatment discontinuation more rigorously than has tended to be done up to now. For example, collecting reasons for study discontinuation such as "lost to follow-up" or "investigator decision" will become completely inadequate. These broad categories do not permit the calculation of important estimands.

Questions about the probability of a patient discontinuing treatment due to tolerability issues on the one hand or due to lack of efficacy on the other can only be answered if detailed reasons for discontinuation are captured.[6]
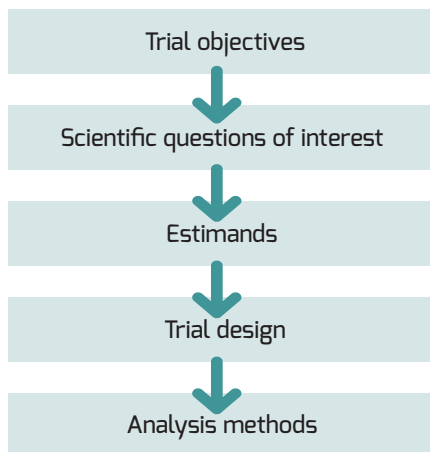

Figure 3. The proposed role of estimands in clinical trial design

Choosing the appropriate estimand for a given trial objective is primarily a medical and clinical question and not a statistical one. Indeed, some prominent statisticians go so far as to proclaim that estimands are not a statistical topic![4] In any case, discussion between medical and statistical experts will be necessary to ensure that the estimands chosen reflect questions of clinical interest and can also be estimated statistically.

In good clinical research it was always the case that researchers started the planning of a trial by defining its objectives. They also chose endpoints and a statistical methodology. However, the potential influence of intercurrent events on the interpretation of the endpoints was rarely considered. Estimands close the gap between the trial objectives and the main estimates by clarifying exactly how intercurrent events will be considered or how the interpretation changes when those events are considered in different ways.

Disclaimers

The opinions expressed in this article are the authors' own and are not necessarily shared by their employers or EMWA.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use [Internet]. Estimands and sensitivity analysis in clinical trials. E9 (R1). Current *Step 2* version. 2017 Jun 16 [cited 2018 Aug 19]. Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/E9-R1EWG_Step2_Guideline_2017_0616.pdf.
2. European Medicines Agency [Internet]. Overview of comments received on draft ICH E9(R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. EMA/744060/2017. 2017 Nov 09 [cited 2018 Aug 19]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Overview_of_comments/2018/04/WC500247968.pdf.
3. ICH E9(R1) Expert Working Group [Internet]. Draft (Step 2) guideline ICH E9(R1) Estimands and Sensitivity Analysis in Clinical Trials (training material). 2018 Jun [cited 2018 Oct 21]. Available from: https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/E9-R1EWG_Modules1–3_Step2_COMPILATION_Training Material_2018_0703.pdf.
4. Hemmings R, Akacha M [Internet]. ICH E9 (R1) Addendum on 'Estimands and Sensitivity Analysis'. Presentation at the Annual Conference of the International Society of Clinical Biostatistics 2017 – Mini Symposia. 2017 [cited 2018 Aug 19]. Available from: http://iscb2017.info/uploadedFiles/ISCB2017.y23bw/fileManager/09%2045%20ROB%20HEMMINGS%20MOUNA%20AKACHA%20 Estimands%20Vigo%20 Final.pdf.
5. Akacha M, Bretz F, Ruberg S. Estimands in clinical trials – broadening the perspective. Stat Med. 2017;36:5–19.
6. Ruberg SJ, Akacha M. Considerations for evaluating treatment effects from randomized clinical trials. Clin Pharmacol Ther. 2017;102:917–23.
7. Akacha M, Kothny W. Estimands: a more strategic approach to study design and analysis. Clin Pharmacol Ther. 2018;102:894–96.
8. Mallinckrodt C, Molenberghs G, Rathmann S. Choosing estimands in clinical trials with missing data. Pharm Stat. 2017;16:29–36.
9. Phillips A, Abellan-Andres J, Soren A, et al. Estimands: discussion points from the PSI estimands and sensitivity expert group. Pharm Stat. 2017;16:6–11.
10. Holzhauer B, Akacha M, Bermann G. Choice of estimand and analysis methods in diabetes trials with rescue medication. Pharm Stat. 2015;14:433–47.
11. Leuchs A-K, Zinserling J, Brandt A, et al. Choosing appropriate estimands in clinical trials. Ther Innov Regul Sci. 2015;49:584–92.
12. Food and Drug Administration [Internet]. Presentations for the July 19, 2011 meeting of the Endocrinologic and Metabolic Drugs Advisory Committee. 2011 [cited 2018 Aug 19]. Available from: https://wayback.archive-it.org/7993/20170405220520/https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/EndocrinologicandMetabolicDrugsAdvisoryCommittee/UCM264312. pdf
13. Bridge H, Schindler TM. The perils of the unknown: missing data in clinical studies. Med Writ. 2018;27:56–9.

Author information

Helen Bridge, DPhil, was previously a university lecturer in German language and literature. She then studied life sciences with statistics at the Open University, and became a regulatory medical writer in 2012. She currently works as an Associate Principal Medical Communications Scientist at AstraZeneca.

Thomas M. Schindler, PhD, studied biology and linguistics, then obtained a PhD in molecular physiology and went on to complete post-doctoral training. Thereafter he went into publishing as a popular science editor and has now gained some 20 years of experience in both medical affairs and regulatory medical writing. He is currently the head of the European regulatory medical writing group at Boehringer Ingelheim Pharma.