Received: 25 July 2019 | Revised: 11 October 2019 | Accepted: 20 October 2019

DOI: 10.1111/bcp.14195

EMA GUIDELINES

Beyond "Intent-to-treat" and "Per protocol": Improving assessment of treatment effects in clinical trials through the specification of an estimand

There is a key problem in randomised clinical trials as outcomes can be distorted due to informative post-randomisation events. This is inadequately addressed by the use of traditional intention-to-treat or per protocol analysis sets and often either ignored or wrongly labelled as missing data. As a consequence, the treatment effects of interest in a clinical trial are not well defined and their estimates might be misinterpreted.

The estimand framework should help all those planning, conducting and analysing clinical trials as well as those interpreting the results to better define, estimate and understand the treatment effects of interest.

This framework is described in the addendum to ICH E9 and addresses precisely this problem. It is relevant for regulatory drug trials and academic-run trials, as well as for trials of nonpharmacological interventions.

1 | PURPOSE AND SCOPE OF THE ICH E9(R1) ADDENDUM (SECTION A.1)

Randomised controlled trials are generally considered to be the optimal experimental design to minimise the biases when estimating treatment effects to investigate the effects of medical interventions.[1]and[2] They are typically conducted to compare the effect of two or more assigned treatment options, one (or more) under investigation and the other one(s) as standard of care or placebo.

Unfortunately, it is not as simple as it looks. Randomisation does protect the treatment comparison from bias due to factors related to the patient or to the environment, known before or at the time of randomisation, but it may not protect from bias arising from events occurring after the initial randomisation that are related to the assigned treatment and influence the outcome of interest. First, such events can preclude the observation of the outcome of interest. One example is the occurrence of the death of a patient in a trial for a progressive, debilitating condition: would the disease have progressed if the patient had stayed alive? Would the treatment under investigation have protected the patient from disease progression? Would death have occurred had the patient been randomised to the group she was not randomised to?

Second, such post-randomisation events can modify the interpretation of the clinical outcome of interest. To study the effect of a disease-modifying treatment in delaying or halting the cognitive decline in early Alzheimer's disease, a long-term trial needs to be conducted, likely running for several years. During follow-up, some patients may permanently discontinue treatment. Some may start on one of the available symptomatic treatments, either whilst still taking the randomised treatment or after discontinuation (Figure 1A). Will either of these events affect the patient score of the cognitive test used as the trial primary endpoint? If so, are we interested in the combined effect of both the investigational disease-modifying treatment and the option to add symptomatic treatments (when available and taken during the trial), or specifically in the effect of the investigational treatment? Another example is a cancer trial where overall survival is considered the most relevant outcome and the medicines under investigation are part of a treatment strategy with different treatments given over time (Figure 1B). Any difference observed between trial arms, or lack thereof, may be influenced by an imbalance in subsequent anticancer therapies taken upon disease progression, including control patients switching to the medicine under investigation. Are we interested in the isolated effect of the investigational treatment or the combined effect of the medicine under investigation and of any subsequent anticancer treatment? Is a comparison between immediate and delayed use of the same medicine meaningful?

These post-randomisation events are well known to clinicians: use of an alternative treatment (such as use of rescue or prohibited medication, or a subsequent line of therapy), changes in background treatments, treatment discontinuation or terminal events (such as death). This is not a comprehensive list as depending on the specific trial many other post-randomisation events might be relevant to consider. But how are they exactly addressed at the design stage of the trial? In common practice some are taken into account in the protocol or in the analysis plan, whereas others are ignored or their handling is not explicit. For example, some are recorded under the general name

of "protocol deviations". This is not informative: a protocol deviation as such does not constitute an interpretable (clinical) event; instead, interest lies in the relevant event that causes patients or investigators to deviate from the protocol, eg sudden worsening of disease that leads to using an additional treatment not allowed in the protocol. More often than not, analysis solutions are defined that set data to missing after such an event has occurred and common missing data analyses are conducted, such as multiple imputation. It is often left to the statistician's judgement whether the estimate of the treatment effect could be biased as a consequence of these post-randomisation events, and how to account for these events in the analysis to reduce potential bias. Thus, these events have typically been handled at the level of analysis sets and statistical analysis, with no explicit discussion on how these choices may alter the research question and impact the clinical interpretation of results.

The current ICH E9[2] guideline describing statistical principles for clinical trials presents us with two main analysis sets that define which patients should be included in the analysis. The "full analysis set", in line with the "intention-to-treat" (ITT) principle, should include all randomised patients and it should be analysed according to treatment

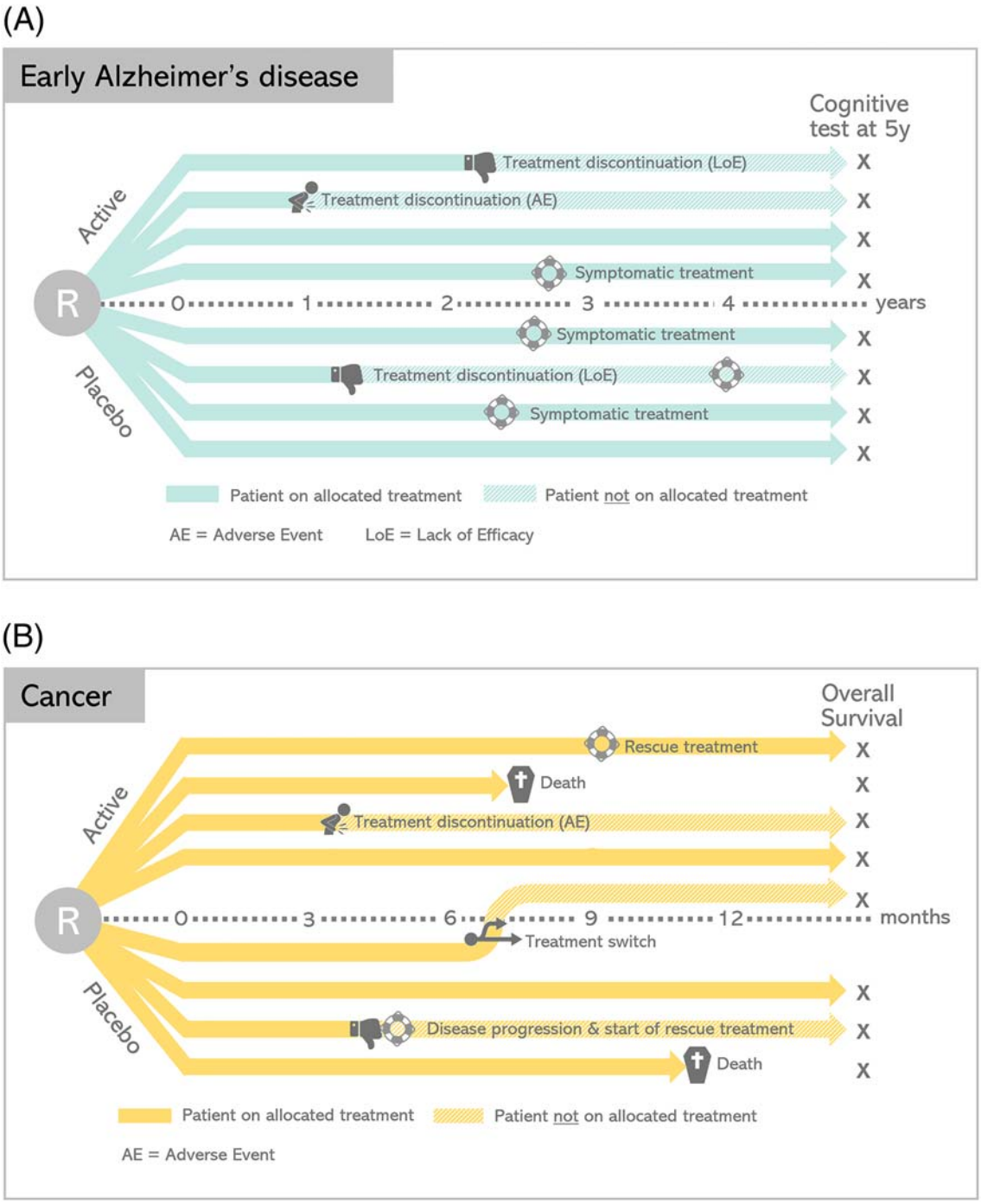

FIGURE 1 Intercurrent events in trials in early Alzheimer's disease (A) and cancer (B)

as randomised. It is generally thought to reflect the effect that would be observed in practice. In contrast, the "per protocol set" restricts the analysis to patients who complied with the procedures detailed in the protocol and it is believed to estimate some "pure" effect of the medicine under investigation.

Once the set of patients for analysis has been defined and data have been collected, values for the outcome of interest may not be available for the analysis: there are "missing data". The subject of analysing data in the presence of missing values for the outcome of interest in medical research has seen extensive development and application in practice since the release of ICH E9 in 1998. Discussion of missing data needs to clearly distinguish two principally different situations. Some data are not collected (observed) but exist; this can happen because patients are not followed up after an alleged intercurrent event, eg due to an adverse event the patient is not willing to continue taking the treatment and decides to leave the trial. Other data do not exist due to an intercurrent event, such as death in case of a longitudinal follow-up trial with symptomatic outcomes. Even though a large armamentarium of methods to address missing data exists, the application of these methods typically does not account explicitly for the underlying cause of the specific data becoming missing. Ideally, estimation requires an approach that explicitly takes these two fundamentally different causes for not observing the data into account to allow for an unbiased estimate of the intended treatment effect.

Surprisingly, even a careful examination of the trial protocol or of the methods section in a scientific publication is often insufficient to find fundamental information: what exact research question does the trial seek to answer? This can be articulated into several subquestions that are not answered explicitly in many trial protocols or scientific publications: what is the medical outcome of interest and when and how is this measured for each individual patient? To which patients should the trial results exactly apply: to all those who could have entered the trial or all those who could tolerate the treatment long enough? Which effect is of interest in a setting where patients can, eg, switch to other treatments or have access to rescue therapy? Is it the effect of starting the new treatment, irrespective of subsequent treatment strategies? Is the study designed and is data collection planned based on these choices (see Section A.4. Impact on Trial Design and Conduct)? Stating the targeted analysis set or principle, eg intention to treat or per protocol, does not fully capture these choices and their consequences for design, data collection and analysis (see Section A.5. Impact on trial analysis). The way trial objectives are usually described often misses out on providing details on a number of corresponding subquestions. An enhanced articulation can greatly improve that and help the interpretation of trial results.

The difficulty in determining causality between treatment and outcome, and estimating the treatment effect without bias based on the per protocol set, can be easily understood. Inclusion of patients in the per protocol set is largely based on post-randomisation events (such as treatment discontinuation due to adverse events) and their behaviour. The patients that comply with the protocol in the experimental arm could be different in their measured and unmeasured baseline characteristics from those in the control arm that comply. The difference observed between arms therefore cannot be attributed to the treatment allocation unless strong assumptions hold in reality.

Problems are also encountered when aiming to follow the ITT principle, where patients "should be followed up, assessed and analysed as members of that group irrespective of their compliance to the planned course of treatment" (ICH E9[2]). Oftentimes a "modified ITT" set is defined, essentially deviating in multiple ways from the ITT principle. Typical deviations include restriction to those patients that received at least one dose of study medication and with at least one post-baseline follow-up assessment. Even under a reasonable definition of the full analysis set, complete follow-up of the primary outcome in the presence of treatment discontinuations is rarely feasible, and methods for handling incomplete datasets necessarily make (implicit) assumptions on the nature of missing values. Methods that are commonly used, such as mixed effects models for longitudinal data, necessarily assume that missingness is not informative after accounting for the covariates and previous measurements in the model (see Section A.5. Impact on trial analysis). Essentially the assumption is that subjects with the same covariates and outcome measurements up to time t will display the same outcome development after time t, irrespective of whether this is observed or not. This is not adequate to address informative post-randomisation events, such as treatment discontinuation that is potentially driven by (unobserved) lack of efficacy or by toxicity. Similar assumptions are made with multiple imputation methods, with the additional complication that the details of such procedures are nearly always underreported.[3]

It is often believed that analyses conducted in line with the ITT principle estimate the effect of treatment assignment in practical clinical scenarios.[4] However, in the case where patients are discontinuing treatment for adverse events (AE) without having their follow-up data collected, the effect that is actually estimated should be interpreted as "the effect of treatment assignment until having to stop randomised treatment due to AE" or "the effect of treatment assignment assuming that patients retain the benefits from treatment after having to stop for AE". These might or (most likely) might not be the effects of interest for the key research question. In the absence of this level of precision in articulating what the target of estimation is, there is the risk of misinterpreting the main results of the clinical trial.

Issues can also arise from application of the ITT principle in the case where full outcome data could be collected. In the case of early Alzheimer's disease referenced above where a disease-modifying treatment is tested, if such a treatment is truly effective, patients assigned to placebo may resort more often to a symptomatic treatment for the duration of the trial. The risk here would be to miss a relevant disease modification effect of the new treatment if the ITT principle is naïvely applied, as this effect could be masked temporarily by the effect of symptomatic medications. An issue arises: the effect targeted by the ITT principle is no longer the effect that is of interest, ie the effect on disease-modification.

2 | A FRAMEWORK TO ALIGN PLANNING, DESIGN, CONDUCT, ANALYSIS AND INTERPRETATION (ADDENDUM SECTION A.2)

The estimand framework, as presented in the ICH E9(R1) Addendum on estimands and sensitivity analysis in clinical trials,[5] aims to provide guidance on articulating the research questions with the above suggested level of precision. It presents a systematic approach to describe the objective of the trial in a more granular manner at the design stage. It does so through drawing attention to the post-randomisation events that are likely to affect the estimation of the treatment effect (labelled "intercurrent events" in the ICH addendum) and reflecting on the best way to address them.

3 | ESTIMANDS (ADDENDUM SECTION A.3)

The estimand framework proposes explicit ways of handling post-randomisation events and presents us with a number of options in addition to the commonly used ITT principle. In the example of disease-modifying treatment, one of these options could be to estimate the effect of the assigned treatment had patients not had access to the symptomatic treatment, or the effect on a particular group of patients who would not require the symptomatic drug. The estimand framework would first shed light on patients starting symptomatic treatment and make us consider these options at an earlier stage, when design aspects can still be tailored. When we consider intercurrent events of a different nature (eg, use of rescue medication or occurrence of an adverse event leading to treatment discontinuation), we may treat them differently in the analysis. For example, for the event of starting symptomatic treatment in early Alzheimer's disease, one could be interested in estimating the treatment effect when patients did not have access to the symptomatic treatment, but for the event of discontinuation due to AE it would not be interesting to pretend that patients had not discontinued the assigned treatment. An alternative is to consider all patients resorting to symptomatic treatments as nonresponders, together with all patients deteriorating by a certain extent (leading to a composite endpoint). In this example, through the estimand framework we would select a separate and potentially different strategy to handle these two distinct intercurrent events, thus clarifying the intended treatment effect explicitly, be it in a trial protocol or a scientific publication.

Only through the precise identification of intercurrent events of interest and the definition of strategies to handle these can the fundamental components of a comparison between clinical interventions[6] be defined precisely: the patient population to which the comparison intends to apply, the interventions, the outcome measure on which the interventions are compared. Subsequently, the study design and analysis methods (see Sections A.4. Impact on trial design and conduct and Section A.5. Impact on trial analysis), and importantly the summary measure of effect, need to be aligned with strategies to deal with the identified intercurrent events. Of course, several strategies can be considered. They will answer distinct research questions and the clinical and/or stakeholder perspective should strongly drive the final choice. This choice goes well beyond simply aiming for an ITT or per protocol analysis. Thus, the estimand framework serves to structure the interdisciplinary discussion to design the trial (see Section A.4. Impact on trial design and conduct) and plan the analysis appropriately (see Section A.5. Impact on trial analysis), and needs substantial clinical and other stakeholder input. It is clearly not intended as a purely statistical or regulatory guideline. The principles described in the addendum to ICH E9 on estimands and sensitivity analysis are considered to be applicable to all clinical trials, including clinical pharmacology and exploratory trials. The choice of strategy could also arguably differ across the stages of drug development, with "hypothetical" effects potentially being of greater interest in earlier rather than later stages. We expect that this will help the generation and communication of more transparent and interpretable information from clinical trials to clinicians and patients.

COMPETING INTERESTS

The authors have no conflicts of interest to declare. The views expressed in this article are the personal views of the authors and may not be understood or quoted as being made onbehalf of or reflecting the position of the agencies or organisations with which the authors are affiliated.

CONTRIBUTORS

F.P. conceived the paper, defined the outline and led the drafting throughout. L.G. and I.A.R. drafted the first version. I.A.R. realised the figures. S.T. and C.B.R reviewed and commented. All authors reviewed and agreed the version published.

Frank Pétavy[1]
Lorenzo Guizzaro[1,2]
Inês Antunes dos Reis[3]
Steven Teerenstra[4,5]
Kit C.B. Roes[4,5]

[1]*European Medicines Agency, Domenico Scarlattilaan 6, Amsterdam, HS, 1083, the Netherlands*
[2]*Medical Statistics, Universita' della Campania Luigi Vanvitelli, Naples, 80138, Italy*
[3]*Medicines and Healthcare products Regulatory Agency, 10 South Colonnade, London, E14 4PU, UK*
[4]*Dutch Medicines Evaluation Board, College ter Beoordeling van Geneesmiddelen, Graadt van Roggenweg 500, Utrecht, AH, 3531, the Netherlands*
[5]*Department of Health Evidence, Section Biostatistics, Radboud University Medical Centre, Geert Grooteplein 21, Nijmegen, GA, 6525, the Netherlands*

Correspondence

Frank Pétavy, European Medicines Agency, Domenico Scarlattilaan 6, 1083. HS Amsterdam, the Netherlands.

Email: frank.petavy@ema.europa.eu

Email: frank.petavy@ema.europa.eu

ORCID

Frank Pétavy https://orcid.org/0000-0003-3668-9922

Lorenzo Guizzaro https://orcid.org/0000-0002-0390-270X

REFERENCES

1. Hadorn DC, Baker D, Hodges JS, Hicks N. Rating the quality of evidence for clinical practice guidelines. *J Clin Epidemiol*. 1996;49(7):749-754.
2. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Statistical Principles for Clinical Trials – E9. Step 5. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf. Published September 1, 1998. .
3. Mackinnon A. The use and reporting of multiple imputation in medical research–a review. *J Intern Med*. 2010;268(6):586-593.
4. Gupta SK. Intention-to-treat concept: a review. *Perspect Clin Res*. 2011;2(3):109-112.
5. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials - E9(R1). Step 4. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf. Published November 20, 2019. .
6. Huang X, Lin J, Demner-Fushman D. Evaluation of PICO as a knowledge representation for clinical questions in AMIA annual symposium proceedings. *American Medical Informatics Association*. 2006;2006:359-363.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.