E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials

For questions regarding this draft document, contact (CDER) Thomas Permutt 301-796-1271 or (CBER) John Scott 240-402-8779.

INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE

ICH HARMONISED GUIDELINE

Estimands and Sensitivity Analysis in Clinical Trials

E9(R1)

Current *Step 1 v*ersion

dated 16 June 2017

At Step 2 of the ICH Process, a consensus draft text or guideline, agreed by the appropriate ICH Expert Working Group, is transmitted by the ICH Assembly to the regulatory authorities of the ICH regions for internal and external consultation, according to national or regional procedures.

STEP 1 Draft Technical Document

Table of Contents

A.1. Purpose and Scope

To properly inform the choices that are made by patients and prescribing physicians, clear descriptions of the effects of a medicine should be available. These descriptions are complicated by the different ways in which each individual patient responds to treatment. Some subjects will tolerate a medicine and adhere to its administration schedule, others will not. Some subjects will require changes in dose of concomitant medication or administration of additional medication, others will not. Multiple ways to quantify treatment effects can be envisaged based on how to take into account, for example, tolerability, adherence and whether or not additional medication is required. Without a precise understanding of the treatment effect that is being described, there is a risk that its magnitude and meaningfulness will be misunderstood.

Confirmatory clinical trials, usually randomised controlled trials, are conducted to quantify the effects of a treatment and to provide evidence of efficacy and safety to support regulatory decision making. Randomised trials are expected to be free from baseline confounding but, in trials as in clinical practice, certain events will occur that complicate the description and interpretation of treatment effects. In this addendum, these are denoted as intercurrent events (see Glossary) and include, among others, use of an alternative treatment (e.g. a rescue medication, a medication prohibited by the protocol or a subsequent line of therapy), discontinuation of treatment, treatment switching and terminal events such as, in some circumstances, death.

Choosing and defining efficacy and safety variables as well as standards for data collection and methods for statistical analysis without first addressing the occurrence of intercurrent events will lead to ambiguity about the treatment effect to be estimated and potential misalignment with trial objectives. The correct order is the reverse. Having clarity in the trial objectives and accounting explicitly for intercurrent events when describing the treatment effect of interest at the planning stage should inform choices about trial design, data collection and statistical analysis.

This addendum presents a structured framework to link trial objectives to a suitable trial design and tools for estimation and hypothesis testing. This framework introduces the concept of an estimand (see Glossary), translating the trial objective into a precise definition of the treatment effect that is to be estimated (Section A.3). It aims to facilitate the dialogue between disciplines involved in clinical trial planning, conduct, analysis and interpretation, as well as between sponsor and regulator, regarding the treatment effects of interest that a clinical trial should address. The statistical analysis, aligned to the estimand, will be associated with assumptions and data limitations, the impact of which can be investigated through sensitivity analysis (see Glossary). This addendum clarifies the definition and the role of sensitivity analysis. References to the original ICH E9 are made using x.y. References within this addendum are made using A.x.y.

This addendum clarifies and extends ICH E9 in a number of respects.

Firstly, ICH E9 introduced the Intention-To-Treat (ITT) principle in connection with the effect of a treatment policy, i.e. the effect of treatment initially assigned at baseline, regardless of adherence to the planned course of treatment, indicating that preservation of randomisation provides a secure foundation for statistical tests. It remains undisputed that randomisation is a cornerstone of controlled clinical trials and that analysis should aim at exploiting the advantages of randomisation to the greatest extent possible. However, the question remains whether understanding the effect of a treatment policy always targets the treatment effect of greatest relevance to regulatory and clinical decision making. The framework outlined in this addendum gives a basis for discussing other treatment effects and some points to consider for the design and analysis of trials to give estimates of these treatment effects that are reliable for decision making.

Secondly, issues considered generally under data handling and missing data (see Glossary) are re-visited. On one hand, intercurrent events such as discontinuation or switching of treatment, or use of rescue medication, may in some circumstances render the later measurements of the variable irrelevant or difficult to interpret even when it can be collected. In the case of death, measurements after a subject dies do not exist. On the other hand, ICH E9 noted the difficulty of fulfilling the ITT principle when clinical trial subjects discontinuing treatment were lost to follow up. This addendum invites consideration of the important distinction between non-adherence with, or withdrawal from, randomised treatment and discontinuation from the trial; also between measurements that exist but have not been collected, and measurements that do not, or cannot, exist. Having clarity in the estimand gives a basis for planning which data need to be collected and hence which data, when not collected, present a missing data problem to be addressed. In turn methods to address the problem presented by missing data can be selected to align with the chosen estimand.

Thirdly, the concept of analysis sets is considered in the proposed framework. Section 5.2 strongly recommends that analysis of superiority trials be based on the full analysis set, defined to be as close as possible to including all randomised subjects. However, trials often include repeated measurements on the same subject. Elimination of some planned measurements on some subjects, perhaps because the measurement is considered irrelevant or difficult to interpret, can have similar consequences to excluding subjects altogether from the full analysis set, i.e. that the initial randomisation is not fully preserved. In addition, a meaningful value of the outcome variable might not exist, as when the subject has died. Section 5.2 does not directly address these issues. Clarity is introduced by carefully defining the treatment effect of interest in a way that determines the population of subjects to be included in the estimation of that treatment effect and the observations from each subject to be included in the analysis considering the occurrence of intercurrent events. The meaning and role of the per-protocol analysis is also re-visited in this addendum; in particular whether the need to explore the impact of protocol violations and deviations can be addressed in a way that is less biased and more interpretable than naïve analysis of the per protocol set.

Finally, the concept of robustness is given expanded discussion under the heading of sensitivity analysis. In particular, a distinction is made between the sensitivity of inference to the particular assumptions of a particular analysis and the sensitivity to the choice of analytic approach more broadly. With precise specification of an agreed estimand and a statistical analysis that is both aligned to the estimand and pre-specified to a level of detail that it can be replicated precisely by a third party, regulatory interest can focus on sensitivity to deviations from assumptions and limitations in the data in respect of a particular analysis.

A.2. A Framework to Align Planning, Design, Conduct, Analysis and Interpretation

To promote coherence and clarity, trial planning should proceed in sequence (Figure 1). Clear trial objectives should be translated into key scientific questions of interest by defining suitable estimands. An estimand defines the target of estimation for a particular trial objective (i.e. "what is to be estimated") through specification of: the population, the variable, the handling of intercurrent events, and the population-level summary for the variable (Section A.3). A suitable method of estimation (i.e. the analytic approach, referred to as the main estimator) can then be selected. The main estimator will be underpinned by certain assumptions. To explore the robustness of inferences from the main estimator to deviations from its underlying assumptions, a sensitivity analysis should be conducted, in form of one or more analyses, targeting the same estimand (Section A.5).

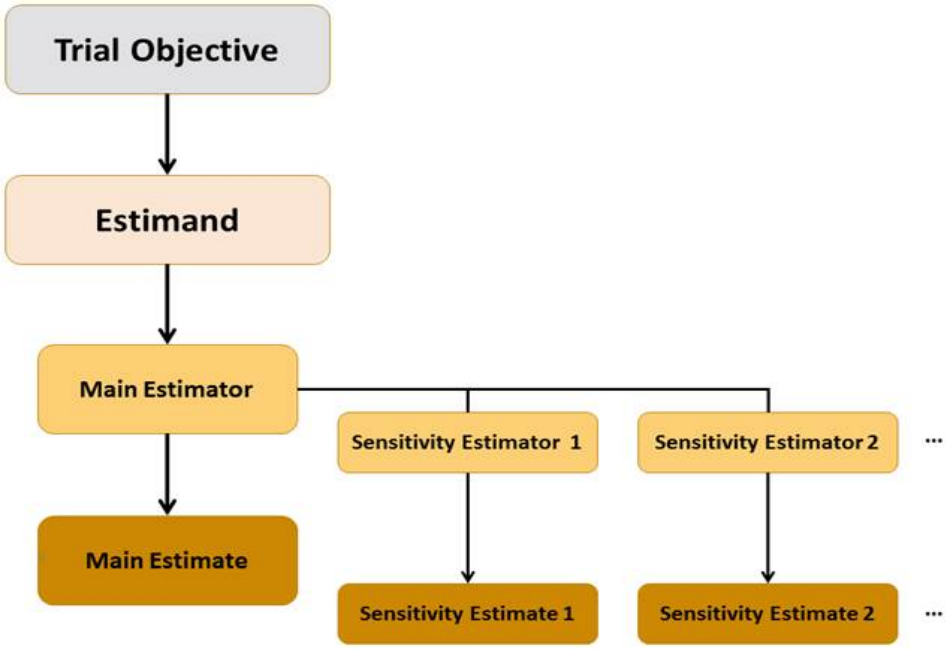

Figure 1: Aligning target of estimation, method of estimation, and sensitivity analysis, for a given trial objective

This framework enables proper trial planning that clearly distinguishes between the target of estimation (trial objective, estimand), the method of estimation (estimator, resulting in an estimate, see Glossary), and a sensitivity analysis. This will assist sponsors in planning trials,

regulators in their reviews, and will enhance the interactions between these parties when discussing the suitability of clinical trial designs, and the interpretation of clinical trial results, to support drug licensing.

In general, it is important to proceed sequentially, and not for the choice of an estimator to determine the estimand, and hence the scientific question that is being addressed.

The specification of appropriate estimands (See A.3.3) will usually be the main determinant for aspects of trial design, conduct (Section A.4) and analysis (Section A.5).

A.3. Estimands

A.3.1. Description

A central question for drug development and licensing is to quantify treatment effects: how the outcome of treatment compares to what would have happened to the same subjects under different treatment conditions (e.g. had they not received the treatment or had they received a different treatment). Intercurrent events need to be considered in the description of a treatment effect on a variable of interest because both the value of the variable and the occurrence of the event may depend on treatment. The definition of a treatment effect, specified through an estimand, should consider whether values of the variable after an intercurrent event are relevant, as well as how to account for the (possibly treatment-related) occurrence or non-occurrence of the event itself.

More formally, an estimand defines in detail what needs to be estimated to address a specific scientific question of interest. A description of an estimand includes four attributes:

A. the population, that is, the patients targeted by the scientific question;
B. the variable (or endpoint), to be obtained for each patient, that is required to address the scientific question;
C. the specification of how to account for intercurrent events to reflect the scientific question of interest.
D. the population-level summary for the variable which provides, as required, a basis for a comparison between treatment conditions

Together these attributes describe the estimand, defining the treatment effect of interest.

In most cases, the target population is reflected by the patients that are eligible to be included in the clinical trial based on the inclusion/exclusion criteria in the protocol. In some cases, a stratum of those patients may be of interest, defined in terms of a potential intercurrent event; for example, the stratum of subjects who would adhere to treatment.

The variable typically consists of measurements taken (e.g., blood pressure measurement), functions thereof (e.g., change from baseline to one year in HbA1c), or quantities related to clinical outcomes (e.g., time of death, times of hospitalisations, number of relapses). The variable may also incorporate intercurrent events such as discontinuation of treatment, for

example when using measurements taken prior to discontinuation (e.g., area under the curve of HbA1c until discontinuation; the number of weeks blood pressure is controlled while on treatment), or composites (e.g., treatment failure defined as non-response or treatment discontinuation).

It is necessary to specify how to account for potential intercurrent events in a way that reflects the scientific question of interest. Intercurrent events can present in multiple forms and can affect the interpretation of the variable. For example, if a subject dies before a planned measurement of blood pressure, the blood pressure will not be observed. If a subject takes rescue medication in addition to treatment, the blood pressure may be observed, but will reflect the combined effect of the treatment and the rescue medication. If a subject discontinues treatment because of toxicity, the blood pressure may be observed but will reflect the lack of effect of the treatment when it is not taken. The set of intercurrent events for consideration will depend on the specific therapeutic setting and trial objective. Taking use of rescue medication as an example, two different specifications include the combined effect of treatment and any intercurrent event (in this case use of rescue medication) and the effect of the treatment in the, potentially hypothetical, absence of the intercurrent event. Section A.3.2 describes different strategies for addressing intercurrent events in constructing an estimand that is best aligned with the corresponding scientific question of interest.

The fourth attribute is the population-level summary measure for the variable, e.g. the mean change from baseline to one year in HbA1c, or the proportion of subjects meeting specified criteria for response. In case of treatment comparisons, the summary measure becomes e.g. the difference in mean change from baseline to one year in HbA1c, or the difference or ratio in the proportion of subjects meeting specified criteria, under two different treatment conditions.

A.3.2. Strategies for Addressing Intercurrent Events

The estimand attributes A through D introduced in Section A.3.1 are inter-related and should not be considered independently. The description of an estimand will not be complete without reflecting how potential intercurrent events are reflected in the scientific question of interest. At least five strategies may be considered. The strategies can be used alone or in combination to address multiple different intercurrent events. Together with the other estimand attributes, the choices made on how to address intercurrent events describe the treatment effect that is targeted. Section A.7 provides illustrations of the use of these five strategies for constructing estimands accounting for one or more intercurrent events.

The relevance of each strategy will depend on the therapeutic and experimental context. In addition it might or might not be possible, in each experimental situation, to derive an estimate for a particular estimand constructed using these strategies that is considered reliable for decision-making. These considerations are addressed in Sections A.3.3, A.3.4, A.4 and A.5. The labels that are presented below are for ease of reference only; an adequate description of the chosen strategy must be used when constructing an estimand.

Treatment policy strategy

The occurrence of the intercurrent event is irrelevant: the value for the variable of interest is used regardless of whether or not the intercurrent event occurs.

For example, when specifying how to account for rescue medication as an intercurrent event, occurrence of the intercurrent event is ignored and the observations on the variable of interest are used. If applied across all types of intercurrent events, this reflects the comparison described in the ICH E9 Glossary (under Intention to Treat Principle) as the effect of a treatment policy.

In general, this strategy cannot be implemented when values for the variable after the intercurrent event do not exist for all subjects. For example, an estimand based on this strategy cannot be constructed with respect to a variable that cannot be measured due to death.

Composite strategy

The occurrence of the intercurrent event is taken to be a component of the variable, i.e. the intercurrent event is integrated with one or more other measures of clinical outcome as the variable of interest.

There are multiple different approaches that can be considered under this label. The requirement to use a rescue medication may provide meaningful information on the effect of a treatment and hence may be incorporated into a variable, with appropriate summary measure, that describes a meaningful treatment effect. For example, the variable might be defined as a composite of no use of rescue medication and a favourable clinical outcome. Alternatively, for a numerical variable, experiencing an intercurrent event might be ascribed an extreme unfavourable value and a suitable summary measure selected. A different approach would be to employ area-under-the curve, reflecting the planned duration of follow-up but based on the values for the variable prior to the intercurrent event.

Sometimes an event being considered as intercurrent is itself the most meaningful variable that can be measured for quantifying the treatment effect of interest. This can be the case with death: the fact that a subject has died may be much more meaningful than observations before death, and observations after death will not exist. For example, in a trial with a primary focus on myocardial infarction, it may not always be possible to ascertain whether a subject who died had, or would have had, a myocardial infarction, but if the variable is defined to be a composite of death or myocardial infarction, this may be completely ascertained.

Hypothetical strategy

A scenario is envisaged in which the intercurrent event would not occur: the value to reflect that scientific question of interest is that which the variable would have taken in the hypothetical scenario defined.

For example, when rescue medication must be made available for ethical reasons, a treatment effect of interest might concern the outcomes if rescue medication had not been available. Analogously, another active treatment might be administered upon failure and subsequent discontinuation of treatment (including treatment switching where the experimental treatment is given to subjects previously randomised to the control arm), but the treatment effect of interest might concern the outcome if the subsequent active treatment had not been administered. In these examples the non-availability of rescue medication and the absence of the other active treatment reflect different hypothetical conditions.

Care is required to precisely describe the hypothetical conditions reflecting the scientific question of interest in the context of the specific trial. For example, the hypothetical condition might usefully address both the use of a rescue medication and adherence to treatment as intercurrent events in order for an estimand to be precisely described.

Principal stratum strategy

The target population might be taken to be the principal stratum (see Glossary) in which an intercurrent event would not occur. For example, the target population of interest might be taken to be the stratum of patients in which failure to adhere to treatment would not occur. In other words, a principal stratum is a subset of the broader population who would not experience the intercurrent event. The scientific question of interest relates to the treatment effect only within that stratum.

Effects in principal strata should be clearly distinguished from any type of subgroup or per-protocol analyses where membership is based on the trial data. Principal stratification (see Glossary) is defined by a patient's potential intercurrent events on both treatments: for example, patients who would adhere to either treatment. It is not possible in general to identify these subjects directly, either in advance of the trial since the occurrence of the intercurrent event cannot be predicted, or based on the data from a randomised controlled trial because each patient will be observed on one treatment only. Membership in a principal stratum must then be inferred, usually imperfectly, from covariates. In contrast, estimation of a treatment effect from any analysis where membership is based on intercurrent events on the assigned treatments is liable to confounding because different subjects will experience different intercurrent events on different treatments.

While on treatment strategy

Response to treatment prior to the occurrence of the intercurrent event is of interest. If a variable is measured repeatedly, its values up to the time of the intercurrent event may be considered to account for the intercurrent event, rather than the value at the same fixed timepoint for all subjects.

For example, subjects with a terminal illness may discontinue a purely symptomatic treatment because they die, yet the success of the treatment can be measured based on the effect on symptoms before death. Alternatively, subjects might discontinue treatment, and in some circumstances it will be of interest to assess the risk of an adverse drug reaction during the period of adherence.

Altogether, five different strategies are considered in this section. It is important to be precise when describing the preferred strategy for handling each intercurrent event. Consider adherence to treatment; it is of utmost importance to distinguish between treatment effects of interest based on (i) the hypothetical scenario of "if all subjects would adhere" from (ii) the stratum of subjects who "would be able to adhere if administered the experimental treatment" and (iii) the effect during adherence.

A.3.3. Construction of Estimands

A.3.3.1. General Considerations

As stated above, in order to unambiguously describe the treatment effect of interest, and to promote the relevance of the treatment effect described to subjects and physicians, intercurrent events need to be considered explicitly in the construction of the estimand. The construction of the estimand should address each intercurrent event that may occur in the clinical trial and that will affect the interpretation of the results of the trial. The description of intercurrent events at the planning stage might in theory reflect very specific details of treatment and follow-up, such as a specific time window for observing a variable. Such specific criteria are not expected to affect interpretation of trial results. It may be impractical to foresee every relevant kind of intercurrent event. Trial reporting should then discuss not only the way unforeseen intercurrent events were handled in the analysis but also the effect on what the chosen analysis estimates. Within the construction of an estimand, different strategies (Section A.3.2, Section A.7) might be selected to address different intercurrent events.

The construction of the estimand(s) in any given clinical trial is a multi-disciplinary undertaking including clinicians, statisticians and other disciplines involved in clinical trial design and conduct. It should be the subject of discussion in a sponsor's interactions with regulators about the objectives and designs for prospective clinical trials. The construction of an estimand should be consequent to the trial objectives and should inform choices relating to data collection and analytic approaches. Avoiding or over-simplifying this process risks misalignment between trial objectives, trial design, data collection and statistical analysis.

An iterative process may be required. The construction of an estimand should be justified considering what is of clinical relevance in the particular therapeutic setting, including the disease under study and the goal of treatment, and the particular experimental setting (Section A.3.3.2). In addition, the adequacy of trial design and statistical methods need to be considered to ensure that an estimate which is reliable for inference can be derived. In particular, the crucial advantage of randomisation in clinical trials should be acknowledged and exploited to the extent possible. Some estimands, in particular those that are estimated using the observed data, can be robustly estimated making few assumptions, whereas other estimands require more specific assumptions that may be more difficult to justify and that may be more sensitive to plausible changes in those assumptions (see Section A.5.1). Where significant issues exist to develop an appropriate trial design or to derive a reliable estimate for a particular estimand, an alternative estimand, trial design and analytic approach would need to be considered.

A.3.3.2. Considerations of Therapeutic and Experimental Context

As indicated above, aspects of the disease setting and the aim of treatment will influence the construction of the estimand. In terms of therapeutic context this might include, respectively, the availability of alternative treatment options and the possibility to monitor individual response to treatment, and whether the treatment is aimed at providing symptom control, modifying the course of the disease or prevention of disease. For example, the goal of a treatment may be control of clinical signs or symptoms in a disease area where multiple alternative treatments exist, with the possibility to tailor the choice of treatment for a patient based on observed response. The use of an alternative treatment (a rescue medication, a medication prohibited by the protocol or a subsequent line of therapy) will likely need to be considered as an intercurrent event. The specification of how to account for intercurrent events to reflect the scientific question of interest might be based on understanding the treatment effect if the alternative treatment was not available, or in the stratum of subjects who can adhere to treatment without needing an alternative. In some circumstances, answers to these questions might be more relevant than e.g. the quantification of the effects of a treatment policy that does not distinguish whether or not a patient has taken an alternative treatment. Such considerations might be of even greater relevance for the intercurrent event of subjects assigned to the control arm switching to treatment. An estimand might be constructed using one of these strategies, providing it is agreed that a robust estimate can be obtained. In other situations, it might be necessary to understand the treatment effect in the context of a treatment policy that exists in clinical practice. For example, the aim of a treatment may be to prevent or delay an adverse clinical outcome (e.g. death). If the treatment is proposed for use in treatment-naïve subjects as part of a treatment policy where subsequent lines of treatment are established, the effect of the treatment policy could be of greater interest. When constructing estimands based on the treatment policy strategy, inference can be complemented by defining an additional estimand and analysis pertaining to the intercurrent event itself; for example, contrasting both the treatment effect on a symptom score and the amount of rescue medication used under each treatment condition.

Estimands based on the treatment policy strategy might also be more generally acceptable to support regulatory decision making, specifically in settings where estimands based on alternative strategies might be considered of greater clinical interest, but main and sensitivity estimators cannot be identified that are agreed to support a reliable estimate or robust inference. An estimand based on the treatment policy strategy might offer the possibility to obtain a reliable estimate of a treatment effect that is still relevant. In this situation, it is recommended to retain those estimands that are considered to be of greater clinical relevance and to present the resulting estimates along with a discussion of the limitations, in terms of trial design or statistical analysis, for that specific approach.

One example for a composite strategy is to replace a continuous variable with a binary variable, in which patients are considered as responders versus non-responders based on a predefined threshold of change in score in the absence of the intercurrent event. This dichotomisation of continuous scores would thus result in a change of the estimand. The clinical relevance and interpretation of the estimand will depend on whether clinically interpretable responder criteria and an appropriate population-level summary (e.g., difference in proportions, odds ratio) are available.

Using the hypothetical strategy, some conditions are likely to be more acceptable for regulatory decision making than others. The hypothetical conditions described must therefore be justified for the quantification of an interpretable treatment effect that is relevant to the use of the medicine in clinical practice. As noted, the question of what the values for the variable of interest would have been if rescue medication had not been available may be an important one, targeting an effect of the treatment under certain conditions rather than a particular treatment policy that includes the use of the rescue medication. In contrast, the question of what the values for the variable of interest would have been under the hypothetical condition that subjects who discontinued treatment because of adverse drug reaction had in fact continued with treatment, might not be justified as being of scientific or regulatory interest. A scientific question of interest based on the effect if all subjects had adhered to treatment is not well-defined without a thorough discussion of the hypothetical conditions under which it is supposed that they would have adhered. Furthermore, the inability to tolerate a treatment in a trial as well as in clinical practice may constitute, in itself, evidence of an inability to achieve a favourable outcome. If the intercurrent event for which a strategy needs to be selected depends not only on, for example, lack of adherence, but also on the reason for the lack of adherence (e.g. due to toxicity), these have to be defined and recorded accurately in the clinical trial.

The experimental situation should also be considered. If patient management (e.g. dose adjustment for intolerance, rescue treatment for inadequate response) under a clinical trial protocol is justified to be different to that which is anticipated in clinical practice, this might be reflected in the construction of the estimand. In particular, the choice of the control arm might influence the manner in which rescue or other concomitant medications are permitted in the trial.

Use of a treatment other than the one assigned will commonly be considered as an intercurrent event. The alternative treatments can be diverse, including rescue medications, medications that are prohibited by the protocol or use of a subsequent line of therapy. Moreover, even rescue medications might be understood in different ways; including use instead of, or in addition to, a chronic treatment on which the subject is experiencing inadequate effect, as an alternative where a subject is not tolerating their assigned treatment, or as a short-term acute treatment to manage a temporary flare in disease symptoms. These examples illustrate the importance of considering the handling of the specific intercurrent event in the context of the particular experimental situation.

The choice of estimands for studies with objectives to demonstrate non-inferiority or equivalence requires careful reflection. In Section 3.3.2 it is stated that such trials are not conservative in nature and the importance of minimising the number of protocol violations and deviations, non-adherence and withdrawals is indicated. In Section 5.2.1, it is described that the result of the Full Analysis Set (FAS) is generally not conservative and that its role in such trials should be considered very seriously. Estimands that are constructed with one or more intercurrent events accounted for using the treatment policy strategy present similar issues for non-inferiority and equivalence trials as those related to the FAS. Responses in both treatment groups will appear more similar following discontinuation of randomised treatment or use of another medication for reasons that are unrelated to the similarity of the initially randomised treatments. Estimands could be constructed to directly address those intercurrent events which can lead to the attenuation of differences between treatment arms (e.g. use of rescue medications and violations from the target population). In this situation, the estimand might target a measure of treatment effect with high sensitivity to detect differences between treatments, if they exist.

A.4. Impact on Trial Design and Conduct

The design of a trial needs to be aligned to the choice of the estimand or estimands that reflect the primary trial objectives and which will form the basis to establish whether those objectives have been met. Specifically, clear definitions for the estimands on which quantification of treatments effects will be based should inform the choices that are made in relation to trial design. If interest lies, for example, in understanding the effect of treatment regardless of whether a particular intercurrent event occurs, a trial in which the variable is collected for all subjects regardless of that event is appropriate. Alternatively, if the estimands that are required to support regulatory decision making do not require the collection of the variable after an intercurrent event, then the benefits of collecting such data for other estimands should be weighed against any complications and potential drawbacks of the collection.

Efforts should be made to collect all data that are relevant to support a statistical analysis aligned to the estimands of interest including important additional estimands. The occurrence

of intercurrent events such as non-adherence, discontinuation of treatment, treatment switching, or use of rescue medication, does not imply that the variable cannot be measured thereafter, unlike for terminal events such as death. Not collecting any data needed to assess an estimand results in a missing data problem for subsequent statistical inference. The validity of statistical analyses may rest upon untestable assumptions and, depending on the proportion of missing data; this may undermine the robustness of the results (Section A.5). A prospective plan to collect informative reasons for why data intended for collection are missing may help to distinguish intercurrent events of interest from residual missing data and thus potentially improve the primary analysis. This may also lead to a more appropriate choice of sensitivity analysis. For example, perhaps a generic "loss to follow up" should correctly be recorded as "treatment discontinuation due to lack of efficacy". Where that has been defined as an intercurrent event of interest, this can be reflected through the chosen strategy to account for that intercurrent event and not as a missing data problem. Measures taken to retain subjects can be implemented, but care should be taken to retain the external validity of the trial to clinical practice. For example, selection of the trial population or use of titration schemes or concomitant medications to mitigate the impact of toxicity might not be suitable if those same measures would not be implemented in clinical practice.

Certain estimands may necessitate, or may benefit from, non-standard trial designs such as run-in or enrichment designs, randomised withdrawal designs, or titration designs. Such alternative designs, however, may require special consideration regarding their implementation and subsequent statistical inference. For example, it might be of interest to try to identify the stratum of subjects who can tolerate a treatment, using a run-in period, in advance of randomising those subjects between treatment and control. Dialogue between regulators and sponsors would need to consider whether the proposed run-in period is appropriate to identify the target population, and whether the choices made for the subsequent trial design (e.g. washout period, randomisation) supports the estimation of the target treatment effect and associated inference. These considerations might limit the use of these trial designs, and use of that particular strategy, in practice.

A precise description of the treatment effects of interest, through specification of strategies to handle intercurrent events, should inform sample size calculations. Where all subjects contribute information to the analysis, and where the impact of intercurrent events and their handling is reflected in the effect size that is targeted and the expected variance, it is not usually necessary to inflate the calculated sample size by the expected proportion of subject withdrawals.

Section 7.2 addresses issues related to summarising data across clinical trials. The need to have consistent definitions for the variables of interest is highlighted and this can be extended to the construction of estimands. Hence in situations when pooling data from across a clinical trial programme is envisaged at the planning stage, a suitable estimand should be constructed, included in the trial protocols, and reflected in the choices made for the designs of the contributing trials. Similar considerations apply to the design of a meta-analysis or the use of external control groups for the interpretation of single-arm trials. A naïve comparison

between data sources, or integration of data from multiple trials without consideration and specification of the estimand that is addressed in each data presentation or statistical analysis, could be misleading and can be considered as a source of bias.

More generally, a trial is likely to have multiple objectives translated into multiple estimands. A trial design that is suitable for one estimand might not be suitable for other estimands of potential importance. Trials with multiple objectives and endpoints might give rise to concerns over multiple testing and in principle these concerns apply equally to the inclusion of multiple estimands. The same approaches employed to address those concerns, in particular the nomination of one or more as primary and others as secondary, can equally be applied to estimands.

A.5. Impact on Trial Analysis

A.5.1. Main Estimation

An estimand for the effect of treatment relative to a control should reflect the outcomes in a group of subjects on the treatment to those in a similar group of subjects on the control, so that the effect of treatment can be isolated from any differences between the groups of subjects on which the comparison is based. For a given estimand an aligned analytic approach, or estimator, should be implemented that is able to provide an estimate on which reliable interpretation can be based. An important consideration for whether a robust estimate will be available is the extent of assumptions that need to be made. Assumptions should be stated explicitly together with the main and sensitivity estimators. Assumptions should be justifiable and implausible assumptions should be avoided. The robustness of the results to the underlying assumptions should be assessed through sensitivity analysis aligned to the estimand (Section A.5.2).

In particular, if there is complete follow-up of subjects regardless of whether or not the intercurrent event occurs, an estimand based on the treatment policy strategy can be estimated with only minimal assumptions. Estimation for an estimand employing this strategy will require stronger and untestable assumptions if measurements are not collected following intercurrent events. Using a composite strategy it may be possible to perform an analysis without need for imputation or modelling of response after an intercurrent event, and the associated assumptions even when the original variable was not completely ascertained. In contrast, the estimation of estimands constructed using a strategy that requires a hypothetical scenario to address an intercurrent event entails careful specification of the hypothetical conditions and will necessarily rely on modelling assumptions that are untestable and need to be investigated through sensitivity analyses. In a randomised trial, estimation of a treatment effect within a principal stratum of the population will be confounded unless the subjects within that stratum can be identified before randomisation. Otherwise, estimation will rely on assumptions, in particular that all relevant confounders have been measured and accounted for. For example, for the stratum of subjects who would be able to adhere to the treatment it is inappropriate to simply compare the observed adherers on the treatment to adherers on

control. These will be systematically different subjects, confounding estimation of the treatment effect. In this case it is essential to account for all important confounders, rather than a small, preconceived set of covariates, though it is difficult to provide assurance against misspecification of the model. For the labelled while-on-treatment strategy, estimation of a treatment effect will require stronger assumptions when the occurrence and timing of an intercurrent event is related to treatment.

Even after defining estimands that address intercurrent events in an appropriate manner, and making efforts to collect the data required for estimation (Section A.4), some data may still be missing. This missing data is distinguished from systematic failure or avoidance in collecting information that are required for estimation. For example, if an estimand based on the treatment policy strategy is constructed, all efforts should be made to retain subjects in the trial and adhere to the schedule of assessments even after discontinuation of assigned therapy. Where those efforts are not successful it becomes necessary to make assumptions about the missing observations, either to predict or impute individual observations or to justify statistical methods based on observed data only. Handling of missing data should be based on plausible assumptions and, where possible, guided by the strategies employed in the description of the estimand. Predictions for a given subject may be based on observed data from that subject (covariates and post-baseline values) and from other similar subjects. Criteria to identify similar subjects might include whether or not the intercurrent event has been assessed (e.g., for subjects who discontinue treatment without further data collected, a prediction model may use data from other subjects who discontinued treatment but for whom data collection has continued rather than from subjects who remained on treatment). Reasonable deviations from the assumptions of these techniques are an important aspect of sensitivity analysis.

A.5.2. Sensitivity Analysis

A.5.2.1. Role of Sensitivity Analysis

Inferences based on a particular estimand should be robust to limitations in the data and deviations from the assumptions used in the statistical model for the main estimator. This robustness is evaluated through a sensitivity analysis.

The statistical assumptions that underpin the main estimator should be documented. One or more analyses, focused on the same estimand, should then be pre-specified to investigate these assumptions with the objective of verifying that the estimate derived from the main estimator is robust to departures from its assumptions. Distinct from this sensitivity analysis, each other analysis that is planned, presented or requested in order to more fully investigate and understand the trial data can be termed supplementary analysis (see Glossary). Each supplementary analysis may refer to a different estimand, or a different estimator to the same estimand. Where the primary estimand(s) of interest is agreed between sponsor and regulator, and the main estimator is pre-specified unambiguously, supplementary analyses should generally be given lower priority than a sensitivity analysis.

A.5.2.2. Choice of Sensitivity Analysis

When planning and conducting a sensitivity analysis, it is recommended not to alter many aspects of the main analysis simultaneously, or else it could be challenging to identify which assumptions, if any, are responsible for any potential differences seen. A more transparent and useful approach is to investigate the impact of changing only one assumption at a time. In addition, a distinction between testable and untestable assumptions may be useful when assessing the interpretation and relevance of different analyses.

Missing data require particular attention in a sensitivity analysis because the assumptions underlying any method may be hard to justify fully and may be impossible to test. Missing data must be defined and considered in respect of a particular estimand. For example, data that were intended to be collected after discontinuation of trial medication to inform an estimand based on the treatment policy strategy are missing if uncollected; however, the same data points might be irrelevant for another strategy, and thus, for the purpose of that second estimand, are not missing if uncollected. Fortunately, relevant types of deviation from assumptions can often be characterized simply. For example, in an analysis of means for continuous outcomes, the original analysis may be biased to the extent that missing and non-missing data for each treatment group differ in their means, and especially when these differences themselves differ across treatment groups. A plausible range of assumed values for these differences should be studied and the robustness of the conclusions assessed. In significance testing, for example, values of the differences for which the treatment effect is or is not statistically significant at a pre-specified level can be plotted in the context of a tipping point analysis. A similar approach can be considered to ascertain values of the differences for which the treatment effect does or does not retain a specific degree of clinical relevance. Similar techniques can be applied to other data structures. For example, proportions of successes or hazards for time-to-event data can be assumed to be different between missing and non-missing data, differentially across treatment groups.

A.5.3. Supplementary Analysis

Interpretation of trial results should focus on the main estimator for each agreed estimand if the corresponding estimate is verified to be robust through the sensitivity analysis.

Supplementary analyses targeting different estimands play a secondary role for interpretation of trial results, though can provide additional insights. For example, an analysis based on the proportion of responders might be helpful for interpretation of a treatment effect that is quantified by difference in mean changes on a continuous scale. Alternatively, different definitions for a responder might be examined to investigate whether the result is robust to that definition. The need for, and utility of, supplementary analyses should be determined for each trial.

Section 5.2.3 indicates that it is usually appropriate to plan for analyses based on both the FAS and the Per-Protocol Set (PPS) so that differences between them can be the subject of explicit discussion and interpretation. Consistent results from analyses based on the FAS and the PPS is indicated as increasing confidence in the trial results. Also in Section 5.2.2 it is described that results based on a PPS might be subject to severe bias. In respect of the framework presented in this addendum, an analysis based on the subset of subjects who adhere to the clinical trial protocol having been assigned to a particular treatment group can be conducted, but does not in itself unambiguously define a treatment effect of interest. As noted above, analysis of the per-protocol data set does not achieve the goal of estimating the effect in adherent subjects because it does not compare similar subjects on different treatments. The role of such an analysis is therefore limited to investigating whether the extent of protocol violations and deviations compromises confidence in the trial results. Some protocol violations and deviations might be addressed as intercurrent events. Where a majority of intercurrent events are handled through the construction of the estimands, the number of remaining protocol violations and deviations will be low and analysis of the PPS might not add additional insights.

A.6. Documenting Estimands and Sensitivity Analysis

Estimands should be defined and explicitly specified in the clinical trial protocol. Having specified those types of intercurrent events that can be foreseen and that would affect the interpretation of the results of the trial, a trial protocol should pre-specify a primary estimand that corresponds to the primary trial objective. Furthermore, the protocol and the analysis plan should pre-specify the main estimator that is aligned with the primary estimand and leads to the primary analysis, together with a suitable sensitivity analysis to explore the robustness under deviations from its assumptions. Estimands for secondary trial objectives (e.g. related to secondary variables) that are likely to support regulatory decisions should be described properly, each with a corresponding main estimator and a suitable sensitivity analysis. Additional trial objectives may be considered for exploratory purposes, leading to additional estimands.

While it is to the benefit of the sponsor to have clarity on what is being estimated, it is not a regulatory requirement to document in detail an estimand for each exploratory question, especially if these are minor variations on primary or secondary estimands in terms of handling intercurrent events. However, where different scientific questions of interest call for materially different estimands, it is recommended that these should be fully documented.

The choice of the primary estimand will usually be the main determinant for aspects of trial design and conduct. Following usual practices, these aspects should be well documented in the trial protocol. If additional estimands are of key interest, these considerations may be extended to support these as needed and should be documented as well. Beyond these aspects, the conventional considerations for trial design, conduct and analysis remain the same. For example, where there is more than one estimand giving rise to potential issues of

multiple testing, the usual considerations for controlling type I error apply and should be described accordingly (Section A.4).

Results from the main, sensitivity and supplementary analyses should be reported systematically in the clinical trial report, specifying whether each analysis was pre-specified, introduced while the trial was still blinded, or performed post hoc. Addressing intercurrent events that were not foreseen at the design stage, or identified during the conduct of the trial should then discuss not only the way intercurrent events were handled in the analysis but the effect on what the chosen analysis estimates and the interpretation of the trial results.

A.7. A Generic Example

In the following, a generic example for a continuous variable is used to illustrate the framework proposed in this addendum. It should not be construed as a regulatory recommendation and should be adapted to the needs of a given clinical trial setting (in particular, but not limited to, when using binary or time to event variables).

A new investigational treatment (Drug X) is considered for subjects with a specific chronic, non-life-threatening disease. Response to treatment is monitored monthly using a continuous measurement. The full effect of Drug X is expected to be seen at four to six months after treatment start. The main scientific question concerns the comparison of Drug X to placebo at month 6, and is best addressed by a randomised clinical trial. Use of placebo in the clinical trial is considered ethical but only if provision is made for subjects to discontinue their treatment and switch to rescue medication due to lack of efficacy. Switch to rescue medication is an intercurrent event, after which it is still possible to collect the variable measurements. This is also the case after other intercurrent events such as discontinuation of treatment due to an adverse event, but not for intercurrent events such as death (considered very unlikely in this setting).

In the unrealistic case where no intercurrent events are expected to occur, the definition of an appropriate estimand is uncontroversial in terms of the following four attributes:

A. Population: defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population for approval;
B. Variable: change from baseline to month six in the designated measurement;
C. Intercurrent event: no intercurrent events to be taken into account;
D. Population-level summary: difference in variable means between treatment conditions.

The estimand is then the difference in means between treatment conditions in the change from baseline to month six in the designated measurement in the targeted patient population.

A design that targets this estimand is a randomised parallel group design where all measurements are collected throughout the trial. Failure to do so would result in missing data. As long as all measurements are collected, an analysis of variance model with

treatment group as a factor is one example for a statistical analysis for this estimand. In case of missing measurements, data need to be predicted based on plausible assumptions that account for the uncertainty due to missing data. For example, missing data may be imputed based on similar subjects who remained in the trial. Similarity may be established based on the same baseline covariates, the same randomised treatment arm, the same measurement history and information on the intercurrent event. Sensitivity analyses should be pre-specified in the trial protocol to assess, for example, the assumptions of the imputation method. Inference can be complemented by including additional supplementary analyses, possibly targeting different estimands, such as contrasting the proportion and timing of rescue switchers between the treatment groups.

Attribute C is labelled as “Intercurrent event” for brevity, referring to the specification of how to account for potential intercurrent events to reflect the scientific question of interest.

A.7.1 One Intercurrent Event

In practice, intercurrent events are expected to occur. For ease of exposition, consider initially the case that only the intercurrent event "switch to rescue medication due to lack of efficacy" is expected to occur. In the following, alternative estimands corresponding to different scientific questions are described, together with high level considerations on trial design, conduct and analysis.

Treatment-policy strategy

A. Population: defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population for approval;
B. Variable: change from baseline to month six in the designated measurement;
C. Intercurrent event: regardless of whether or not switching to rescue medication had occurred;
D. Population-level summary: difference in variable means between treatment conditions.

In this specific example the estimand described by the treatment-policy strategy is the effect of "Drug X + rescue medication as needed" versus "placebo + rescue medication as needed" on the variable measurement. Thus, dependent on the proportion of rescue medication switchers in both treatment arms, this estimand captures a mixture of the effects of treatment and rescue medication. Also, this estimand does not capture that switching to rescue medication is driven by the unfavourable event of "lack of efficacy".

The estimand is then the difference in means between treatment conditions in the change from baseline to month six in the designated measurement in the targeted patient population, regardless of whether or not switching to rescue medication had occurred.

A similar sentence can be constructed for each of the examples below, also integrating the specification for how the intercurrent events are handled.

A design that targets this estimand is a randomised parallel group design where all measurements regardless of switching to rescue medication are collected throughout the trial.

As long as all measurements are collected, an analysis of variance model with treatment group as a factor is one example for a statistical analysis for this estimand. In case of missing measurements, data need to be predicted based on plausible assumptions that account for the uncertainty due to missing data. For example, missing data may be imputed based on similar subjects who remained in the trial. Similarity may be established based on the same baseline covariates, the same randomised treatment arm, the same measurement history and information on the intercurrent event. Sensitivity analyses should be pre-specified in the trial protocol to assess, for example, the assumptions of the imputation method. Inference can be complemented by including additional supplementary analyses, possibly targeting different estimands, such as contrasting the proportion and timing of rescue switchers between the

757 treatment groups. Another estimand of interest could be constructed to address a scientific
758 question on the use of rescue medication.

Composite strategy

A. Population: defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population for approval;
B. Variable: binary response variable indicating a successful response at month six if the change from baseline to month six in the designated measurement is above a pre-specified threshold, and no switching to rescue medication occurred;
C. Intercurrent event: the intercurrent event is captured through the variable definition;
D. Population-level summary: difference in response proportions between treatment conditions.

The estimand described by the composite strategy no longer assesses the treatment effect only in terms of the variable measurements at month six. Rather, the treatment effect is established based on a composite variable which combines a clinically meaningful dichotomous change in the variable measurement with the intercurrent event of "switching to rescue". As switching to rescue medication is based on lack of efficacy, this estimand acknowledges that intake of rescue medication is an unfavourable outcome.

A design that targets this estimand is a randomised parallel group design. There would be no need to collect measurements after switching to rescue medication, unless there is interest in alternative trial objectives that would require such data (e.g. to collect safety information even after the intercurrent event). In this example, data that could have been collected after the use of rescue medication is not regarded as missing as they are not of interest for estimating the targeted estimand.

As long as all measurements to establish the response status are collected, a logistic regression is one example for a statistical analysis for this estimand. In case of missing data, i.e. prior to the assessment point without an intercurrent event having occurred, the response status needs to be imputed based on plausible assumptions that account for the uncertainty due to missing data. For example, missing data may be imputed based on similar subjects who remained in the trial. Similarity may be established based on the same baseline covariates, the same randomised treatment and the same measurement history. Sensitivity analyses should be pre-specified in the trial protocol to assess, for example, the assumptions of the imputation method. Inference can be complemented by including additional supplementary analyses targeting the separate components of this composite estimand, such as changing the threshold in the variable definition, leading to a different estimand.

Hypothetical strategy

A. Population: defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population for approval;
B. Variable: change from baseline to month six in the designated measurement;
C. Intercurrent event: had rescue medication not been made available to subjects prior to month six;
D. Population-level summary: difference in variable means between treatment conditions.

The estimand described by the hypothetical strategy addresses the treatment effect in an alternative, hypothetical setting where rescue medication was not available to subjects. Conducting a clinical trial to target this scientific question directly may not be ethically justifiable.

A design that targets the hypothetical estimand is a randomised parallel group design. There would be no need to collect measurements after switching to rescue medication, unless there is interest in alternative trial objectives that would require such data (e.g. to collect safety information even after the intercurrent event). In this example, data that could have been collected after the use of rescue medication is not regarded as missing as they are not of interest for estimating the targeted estimand.

A statistical analysis for this estimand will rest on assumptions about the measurements that would have been observed under the hypothetical setting where rescue medication was not available to subjects. Generally, the assumptions needed for such predictions cannot be verified based on the observed data so that a sensitivity analysis will be necessary to assess the robustness of conclusions. A discussion on the plausibility of the assumptions will be warranted to give sufficient credibility to these assumptions, and as a consequence the estimation of the treatment effect. Inference can be complemented by including additional supplementary analyses, possibly targeting different estimands, such as contrasting the proportion and timing of rescue switchers between the treatment groups.

Principal stratum strategy

A. Population: defined through subjects who would not require rescue medication over a period of six months regardless of treatment assignment, within the targeted population defined by inclusion/exclusion criteria;
B. Variable: change from baseline to month six in the designated measurement;
C. Intercurrent event: the intercurrent event is captured through the population definition;
D. Population-level summary: difference in variable means between treatment conditions.

The estimand described by the principal stratum strategy assesses the effect of the initially randomised treatments in the stratum of the population who would not require rescue medication over a period of six months regardless of which treatment arm they were randomised to.

One complication with this estimand is that, in practice, it is difficult to identify the members of this population in advance. Thus, in practice one may have to employ non-standard designs to target patients that would not require rescue medication over a period of six months, such as enrichment designs as well as run-in and randomised withdrawal designs.

A statistical analysis for this estimand is straightforward as long as only subjects who would not require rescue medication over a period of six months had been randomised, and they were followed for the entire trial duration. As noted above, however, it is generally difficult

to identify the members of this population in advance. If the targeted population cannot be identified, then a suitable analysis cannot be achieved by restricting the analysis to those subjects who did not switch to rescue medication: this could exclude systematically different subjects on the different assigned treatments, so that the treatment effect would be confounded with patient characteristics that affect the subjects' propensity to switch to rescue medication. An appropriate analysis needs to account for this confounding. In addition, an assessment of the robustness of conclusions to the assumptions made is necessary using appropriate sensitivity analyses. Inference can be complemented by including additional supplementary analyses, possibly targeting different estimands, such as contrasting the proportion and timing of rescue switchers between the treatment conditions.

While on treatment strategy

A. Population: defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population for approval;
B. Variable: average of the designated measurements while on randomised treatment;
C. Intercurrent event: the intercurrent event is captured through the variable definition;
D. Population-level summary: difference in variable means between treatment conditions.

This estimand assesses the average treatment effect on the variable measurement. The variable chosen here averages the outcomes while being on treatment, i.e. before switch to rescue medication.

A design that targets this estimand is a randomised parallel group design. There would be no need to collect measurements after switching to rescue medication, unless there is interest in alternative trial objectives that would require such data (e.g. an alternative estimand that requires those data, or to collect safety information even after the intercurrent event). In this example, data that could have been collected after the use of rescue medication are not regarded as missing as they are not of interest for estimating the targeted estimand.

As long as all measurements while on the randomised treatments are collected, an analysis of variance model with treatment group as a factor is an appropriate statistical analysis for this estimand. In case of intermittent missing measurements, data need to be interpolated based on plausible assumptions that account for the uncertainty due to missing data. Sensitivity analyses should be pre-specified in the trial protocol to assess, for example, the assumptions of the interpolation method. Inference can be complemented by including additional supplementary analyses, possibly targeting different estimands, such as considering alternative choices for the variable definition by focussing on the last measurement while being on treatment, leading to different estimands.

A.7.2. Two Intercurrent Events

The generic example is now extended to situations where two types of intercurrent events may occur, namely "switch to rescue medication" and "discontinuation of treatment due to an adverse event". The definition of a clinically meaningful estimand needs to encompass all intercurrent events that are likely to occur and are clinically relevant in a given clinical trial setting, to the extent that the description of the treatment effect being targeted cannot be fully understood without inclusion of the intercurrent event in the estimand. The same holds for choices made about the design, conduct and statistical analysis. Considering the five strategies discussed above, all possible combinations of strategies for two types of intercurrent events can be considered, although not all combinations will be clinically relevant. For ease of exposition, only two different estimand strategies are described in the following, together with high level considerations on trial design, conduct and analysis.

Treatment-policy strategy to account for both intercurrent events

A. Population: defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population for approval;
B. Variable: change from baseline to month six in the designated measurement;
C. Intercurrent events: regardless of switching to rescue medication and regardless of treatment discontinuation due to an adverse event;
D. Population-level summary: difference in variable means between treatment conditions.

This estimand targets the treatment-policy effect of treatment initiation on the variable measurement. This estimand accounts neither for rescue medication initiation nor for treatment discontinuation due to an adverse event. In particular, it does not capture that switching to rescue medication and adverse events are unfavourable outcomes.

A design that targets this estimand is a randomised parallel group design where all measurements regardless of switching to rescue medication and treatment discontinuation due to adverse events are collected throughout the trial.

As long as all measurements are collected, an analysis of variance model with treatment group as a factor is an appropriate statistical analysis for this estimand. In case of missing measurements, data need to be predicted based on plausible assumptions that account for the uncertainty due to missing data. For example, missing data may be imputed based on similar subjects who remained in the trial. Similarity may be established based on the same baseline covariates, the same randomised treatment arm, the same measurement history and information on the intercurrent events. Sensitivity analyses should be pre-specified in the trial protocol to assess, for example, the assumptions of the imputation method. Inference can be complemented by including additional supplementary analyses, possibly targeting different estimands, such as contrasting the proportion and timing of rescue switchers and treatment discontinuations due to adverse events between the treatment groups.

Combination of Hypothetical strategy and Treatment-policy strategy to account for the two intercurrent events

A. Population: defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population for approval;
B. Variable: change from baseline to month six in the designated measurement;
C. Intercurrent events: had rescue medication not been made available to subjects prior to month six and regardless of study treatment discontinuation due to an adverse event;
D. Population-level summary: difference in variable means between treatment conditions.

This estimand combines two different strategies to account for the two types of intercurrent events. It employs a hypothetical strategy to address switching to rescue medication and a treatment-policy strategy to address treatment discontinuation due to an adverse event. Such an estimand may be of interest and easily interpretable in settings where the pharmacological

effect is targeted but withholding rescue medication is not ethical and where subjects remain untreated after treatment discontinuation due to an adverse event.

A design that targets this estimand is a randomised parallel group design where all measurements regardless of treatment discontinuation due to an adverse event are collected throughout the trial. There would be no need to collect measurements after switching to rescue medication, unless there is interest in alternative trial objectives that would require such data. In this example, data that could have been collected after the use of rescue medication are not regarded as missing.

A statistical analysis for this estimand needs to account for both intercurrent events:

- Switching to rescue medication: Interest lies in the effect had rescue medication not been made available to subjects prior to month six. As measurements under this scenario cannot be directly observed, assumptions about the measurements that would have been observed under this hypothetical setting need to be made.
- Study treatment discontinuation due to an adverse event: Interest lies in the effect regardless of this intercurrent event. Thus, all measurements regardless of this intercurrent event need to be included in the analysis. In case of missing measurements, data need to be predicted based on plausible assumptions while accounting for the added uncertainty due to missing data. For example, missing data may be imputed based on similar subjects who remained in the trial. Similarity may be established based on the same baseline covariates, the same randomised treatment arm, the same measurement history and information on the intercurrent event, e.g. timing.

Once the individual predictions are made in line with the observed intercurrent events and the estimand of interest, a statistical analysis using, for example, an analysis of variance model based on all randomised subjects is appropriate. In case of missing measurements, data need to be predicted based on plausible assumptions that account for the uncertainty due to missing data. For example, missing data may be imputed based on similar subjects who remained in the trial. Similarity may be established based on the same baseline covariates, the same randomised treatment arm, the same measurement history and information on the intercurrent events. Sensitivity analyses should be pre-specified in the trial protocol to assess, for example, the assumptions of the imputation method. Inference can be complemented by including additional supplementary analyses, possibly targeting different estimands, such as contrasting the proportion and timing of rescue switchers and treatment discontinuations due to adverse events between the treatment groups.

Glossary

Estimand:
Is the target of estimation to address the scientific question of interest posed by the trial objective. Attributes of an estimand include the population of interest, the variable (or endpoint) of interest, the specification of how intercurrent events are reflected in the scientific question of interest, and the population-level summary for the variable.

Estimate:
Is the numerical value computed by an estimator based on the observed clinical trial data.

Estimator:
Is the analytic approach to compute an estimate from observed clinical trial data.

Intercurrent Events:
Events that occur after treatment initiation and either preclude observation of the variable or affect its interpretation.

Missing Data:
Data that would be meaningful for the analysis of a given estimand but were not collected. They should be distinguished from data that do not exist or data that are not considered meaningful because of an intercurrent event.

Principal Stratification:
Is the classification of subjects according to the potential occurrence of an intercurrent event on all treatments. With two treatments, there are four principal strata with respect to a given intercurrent event: subjects who would not experience the event on either treatment, subjects who would experience the event on treatment A but not B, subjects who would experience the event on treatment B but not A, and subjects who would experience the event on both treatments.

Principal Stratum:
Is used in this document to refer to any of the strata (or combination of strata) defined by principal stratification.

Sensitivity Analysis:
Is a series of analyses targeting the same estimand, with differing assumptions to explore the robustness of inferences from the main estimator to deviations from its underlying modelling assumptions and limitations in the data.

Supplementary Analysis:
Is a general description for analyses that are conducted in addition to the main and sensitivity analysis to provide additional insights into the understanding of the treatment effect. The term describes a broader class of analyses than sensitivity analyses.