E9(R1) STATISTICAL PRINCIPLES FOR CLINICAL TRIALS: ADDENDUM: ESTIMANDS AND SENSITIVITY ANALYSIS IN CLINICAL TRIALS

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2021
ICH

Revision 1

E9(R1) STATISTICAL PRINCIPLES FOR CLINICAL TRIALS: ADDENDUM: ESTIMANDS AND SENSITIVITY ANALYSIS IN CLINICAL TRIALS

Guidance for Industry

Additional copies are available from:
Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353
Email: druginfo@fda.hhs.gov
https://www.fda.gov/drugs/guidance-compliance-regulatory-information/guidances-drugs
and/or
Office of Communication, Outreach and Development
Center for Biologics Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 71, Room 3128
Silver Spring, MD 20993-0002
Phone: 800-835-4709 or 240-402-8010
Email: ocod@fda.hhs.gov
https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-biologics/biologics-guidances

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2021
ICH

Revision 1

FOREWORD

The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) has the mission of achieving greater regulatory harmonization worldwide to ensure that safe, effective, and high-quality medicines are developed, registered, and maintained in the most resource-efficient manner. By harmonizing the regulatory expectations in regions around the world, ICH guidelines have substantially reduced duplicative clinical studies, prevented unnecessary animal studies, standardized safety reporting and marketing application submissions, and contributed to many other improvements in the quality of global drug development and manufacturing and the products available to patients.

ICH is a consensus-driven process that involves technical experts from regulatory authorities and industry parties in detailed technical and science-based harmonization work that results in the development of ICH guidelines. The commitment to consistent adoption of these consensus-based guidelines by regulators around the globe is critical to realizing the benefits of safe, effective, and high-quality medicines for patients as well as for industry. As a Founding Regulatory Member of ICH, the Food and Drug Administration (FDA) plays a major role in the development of each of the ICH guidelines, which FDA then adopts and issues as guidance to industry.

TABLE OF CONTENTS

Contains Nonbinding Recommendations

E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials Guidance for Industry[1]

This guidance represents the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA office responsible for this guidance as listed on the title page.

I. PURPOSE AND SCOPE (A.1)[2]

To properly inform decision-making by pharmaceutical companies, regulators, patients, physicians, and other stakeholders, clear descriptions of the benefits and risks of a treatment (medicine) for a given medical condition should be made available. Without such clarity, there is a concern that the reported *treatment effect* will be misunderstood. This addendum presents a structured framework to strengthen the dialogue between disciplines involved in the formulation of clinical trial objectives, design, conduct, analysis and interpretation, as well as between sponsor and regulator regarding the treatment effect or effects of interest that a clinical trial should address.

Precision in describing a treatment effect of interest is facilitated by constructing the *estimand* (see Glossary; section III (A.3)) corresponding to a clinical question of interest. Clarity calls for a thoughtful envisioning of *intercurrent events* (see Glossary; section III.A (A.3.1)) such as discontinuation of assigned treatment, use of an additional or alternative treatment, and terminal events such as death. The description of an estimand should reflect the clinical question of interest in respect to these intercurrent events, and this addendum introduces strategies to reflect different questions of interest that might be posed. The choice of strategies can influence how more conventional attributes of a trial are reflected when describing the clinical question, for example, the treatments, population, or the variable (endpoint) of interest.

The statistical analysis of clinical trial data should be aligned to the estimand. This addendum clarifies the role of *sensitivity analysis* (see Glossary) to explore robustness of conclusions from the main statistical analysis.

[1] This guidance was developed within the Expert Working Group (Efficacy) of the International Council on Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) and has been subject to consultation by the regulatory parties, in accordance with the ICH process. This document has been endorsed by the ICH Steering Committee at *Step 4* of the ICH process, November 2019. At *Step 4* of the process, the final draft is recommended for adoption to the regulatory bodies of the ICH regions.

[2] The numbers in parentheses reflect the organizational breakdown of the document endorsed by the ICH Steering Committee at *Step 4* of the ICH process, November 2019.

Throughout the addendum, references to the main body of the ICH guidance for industry *E9 Statistical Principles for Clinical Trials* (September 1998)[3] are made using x.y in parentheses. References within this addendum are made using A.x.y in parentheses.

This addendum clarifies and extends ICH E9 in respect to the following topics.

First, ICH E9 introduced the Intention-To-Treat (ITT) principle in connection with the effect of a treatment policy in a randomized controlled trial, whereby subjects are followed, assessed, and analyzed irrespective of their compliance to the planned course of treatment, indicating that preservation of randomization provides a secure foundation for statistical tests. Three consequences arising from the ITT principle can be distinguished: (1) that the trial analysis should include all subjects relevant for the research question; (2) that subjects should be included in the analysis as randomized; and (3) as taken directly from the definition of the ITT principle (see ICH E9 Glossary), that subjects should be followed up and assessed regardless of adherence to the planned course of treatment and that those assessments should be used in the analysis. It remains undisputed that randomization is a cornerstone of controlled clinical trials and that analysis should aim at exploiting the advantages of randomization to the greatest extent possible. However, the question remains whether estimating an effect in accordance with the ITT principle always represents the treatment effect of greatest relevance to regulatory and clinical decision-making. The framework outlined in this addendum gives a basis for describing different 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.

Second, issues considered generally under data handling and *missing data* (see Glossary) are revisited. Two important distinctions are made: (1) the addendum distinguishes discontinuation of randomized treatment from study withdrawal and (2) the addendum highlights the distinct consequences of different intercurrent events. Discontinuation of randomized treatment represents an intercurrent event to be addressed in the precise specification of the trial objective through the estimand. Study withdrawal gives rise to missing data to be addressed in the statistical analysis. Consider, for example, a subject switching treatments in an oncology trial, and a subject for whom no outcome event can be observed because the trial is completed. Switching treatment represents an intercurrent event, and the clinical question of interest in respect to that event should be clear. Trial completion before an outcome is observed is administrative censoring that should be addressed as a missing data problem in the statistical analysis. Having clarity in the estimand gives a basis for planning which data should be collected and hence which data, when not collected, present a missing data problem to be addressed in the statistical analysis. In turn, methods to address the problem presented by missing data can be selected to align with the estimand. In addition, regarding the distinct consequences of different intercurrent events, events such as discontinuation of treatment, switching between treatments, or use of an additional medication may render the later measurements of the variable irrelevant or difficult to interpret even when they can be collected. Measurements after a subject dies do not exist.

[3] We update guidances periodically. For the most recent version of a guidance, check the FDA guidance web page at https://www.fda.gov/regulatory-information/search-fda-guidance-documents.

Third, issues related to the concept of analysis sets are considered in the framework. Section V.B (5.2) of ICH E9 strongly recommends that analysis of superiority trials be based on the full analysis set, defined to be as close as possible to including all randomized subjects. However, trials often include repeated measurements on the same subject. Eliminating 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 randomization is not fully preserved. A consequence of this not fully preserving the initial randomization is that the theoretical benefits that randomization confers on testing hypotheses about treatment effects and the practical benefits of balancing confounding factors at baseline can be diminished. In addition, a meaningful value of the outcome variable might not exist, such as when the subject dies. Section V.B (5.2) does not directly address these issues. Clarity is introduced by carefully defining the treatment effect of interest in a way that determines both 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 an analysis of the per protocol set is also revisited in this addendum, in particular, whether 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 (see section I.B (1.2)) is given expanded discussion under the heading of sensitivity analysis. A distinction is made between the sensitivity of inference to the assumptions of a chosen method of analysis and the sensitivity to the choice of analytic approach more broadly. With precise specification of an agreed-upon estimand and a method of analysis that is both aligned to the estimand and prespecified 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 to a particular analysis.

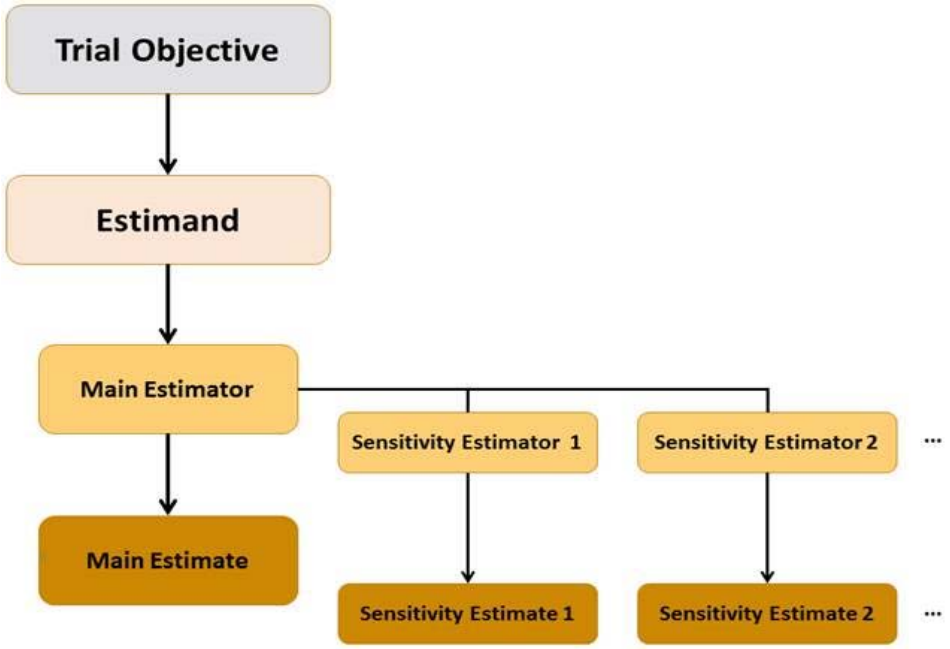
The principles outlined in this addendum are relevant whenever a treatment effect is estimated or a hypothesis related to a treatment effect is tested, whether related to efficacy or safety. Although the main focus is on randomized clinical trials, the principles are also applicable for single-arm trials and observational studies. The framework applies to any data type, including longitudinal, time-to-first event, and recurrent event data. Regulatory interest in the application of the principles outlined will be greater for confirmatory clinical trials and, when used to generate confirmatory conclusions, for data integrated across trials.

The contents of this document do not have the force and effect of law and are not meant to bind the public in any way, unless specifically incorporated into a contract. This document is intended only to provide clarity to the public regarding existing requirements under the law. FDA guidance documents, including this guidance, should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. A FRAMEWORK TO ALIGN PLANNING, DESIGN, CONDUCT, ANALYSIS AND INTERPRETATION (A.2)

Trial planning should proceed in sequence (see Figure 1). Clear trial objectives should be translated into key clinical 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*, see section III (A.3) of this guidance). A suitable method of estimation (i.e., the analytic approach, referred to as the main *estimator*, see Glossary) can then be selected (see section V.A (A.5.1)). Certain assumptions will underpin the main estimator. To explore the robustness of inferences from the main estimator to deviations from its underlying assumptions, a sensitivity analysis should be conducted in the form of one or more analyses, targeting the same estimand (see section V.B (A.5.2)).

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), the numerical result (*estimate*, see Glossary), and a sensitivity analysis. This framework will assist sponsors in planning trials, assist regulators in their reviews, and enhance the interactions between these parties when discussing the suitability of clinical trial designs and, the interpretation of clinical trial results.

The specification of appropriate estimands (see section III (A.3)) will usually be the main determinant for aspects of trial design, conduct (see section IV (A.4)) and analysis (see section V (A.5)).

III. ESTIMANDS (A.3)

Central questions for drug development and licensing are to establish the existence, and to estimate the magnitude, of treatment effects: how the outcome of treatment compares to what would have happened to the same subjects under alternative treatment (i.e., had they not received the treatment, or had they received a different treatment). An estimand is a precise description of the treatment effect reflecting the clinical question posed by a given clinical trial objective. It summarizes at a population level what the outcomes would be in the same patients under different treatment conditions being compared. The targets of estimation should be defined in advance of a clinical trial. Once defined, a trial can be designed to enable reliable estimation of the targeted treatment effect.

The description of an estimand involves precise specifications of certain attributes, which should be developed based not only on clinical considerations but also on how intercurrent events are reflected in the clinical question of interest. Section III.A (A.3.1) introduces intercurrent events. Section III.B (A.3.2) introduces strategies to describe the question of interest in respect to intercurrent events. Section III.C (A.3.3) describes the attributes of an estimand, and section IV.D (A.3.4) gives considerations for its construction. It is critically important to understand the differences between the strategies and to precisely articulate which are used in constructing the estimand.

A. Intercurrent Events to be Reflected in the Clinical Question of Interest (A.3.1)

Intercurrent events are events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest. It is important to address intercurrent events when describing the clinical question of interest to precisely define the treatment effect that is to be estimated.

Intercurrent events should be considered in the description of a treatment effect, because measurements of the variable can be influenced by the intercurrent event and the occurrence of the intercurrent event may depend on treatment. For example, two patients might be exposed initially to the same treatment and provide the same measure of outcome, but if one patient has received additional medication, the information that the two measures give about the treatment differs between the two patients. Furthermore, whether a patient needs to take additional medication, and whether a patient can continue taking treatment, may depend on the treatment to which the patient is exposed. Unlike missing data, intercurrent events should not be thought of as a drawback to be avoided in clinical trials. Discontinuation of prescribed treatment, use of additional medication, and other such events may occur in clinical practice as they do in clinical trials, and their occurrence should be considered explicitly when defining the clinical question of interest.

Examples of intercurrent events that can affect interpretation of the measurements include discontinuation of assigned treatment and use of an additional or alternative therapy. Use of an additional or alternative therapy can take multiple forms, including change to background or concomitant therapy and switching between treatments of interest. Examples of intercurrent events that would affect the existence of the measurements include terminal events such as death and leg amputation (when assessing symptoms of diabetic foot ulcers), when these events are not part of the variable itself. Certain clinical events can also be intercurrent events, when their occurrence, or nonoccurrence, defines a principal stratum of interest (see section III.B (A.3.2)). Examples include tumor shrinkage defining objective response when assessing a treatment effect on duration of response in oncology and occurrence of infection when assessing a treatment effect on severity of infections occurring after vaccination of initially uninfected subjects.

An intercurrent event might be identified solely by the event itself, such as discontinuation of treatment, or might be more granular. For example, the reason for the event might be specified, such as discontinuation of treatment due to toxicity, or due to lack of efficacy; the event might be characterized by a certain magnitude or degree, such as the use of additional medication exceeding a specified duration or dose; or the timing of the event might be specified, perhaps in relation to its proximity to the assessment of the variable. Some events will affect interpretation of the outcome measurements indefinitely, such as discontinuation of treatment, while others will affect interpretation only temporarily, such as short-term use of additional treatment. Indeed, additional, or alternative treatments can be diverse, either replacing or supplementing a treatment on which the subject is experiencing inadequate benefit, as an alternative when a subject is not tolerating the assigned treatment, or as a short-term acute treatment to manage a temporary flare in disease symptoms. In a clinical trial, additional or alternative treatments are often identified as, for example, background treatment, rescue medication, prohibited medication, distinguishing their different roles and allowing them to be considered separately. The additional granularity, identifying different intercurrent events, is necessary if different strategies are to be used. If the intercurrent event for which a strategy should be selected depends not only on, for example, failure to continue with treatment, but also on the reason, magnitude, or timing associated with that failure, this additional information should be defined and recorded accurately in the clinical trial. The description of intercurrent events might in theory reflect very specific details of treatment and follow-up, such as a single missed dose of a chronic treatment or a dose taken at the wrong time of day. If such specific criteria are not expected to affect interpretation of the variable, they would not need to be addressed as intercurrent events.

As indicated above, intercurrent events should be considered when constructing the estimand. Because the estimand should be defined in advance of trial design, neither study withdrawal nor other reasons for missing data (e.g., administrative censoring in trials with survival outcomes) are in themselves intercurrent events. Subjects who withdraw from the trial may have experienced an intercurrent event before withdrawal.

B. Strategies for Addressing Intercurrent Events when Defining the Clinical Question of Interest (A.3.2)

Descriptions of various strategies are listed below, each reflecting a different clinical question of interest in respect to a particular intercurrent event. Whether or not the naming convention is

used, the choices of strategy should be unambiguously clear once the estimand is constructed. It is not necessary to use the same strategy to address all intercurrent events. Indeed, different strategies will often be used to reflect the clinical question of interest in respect to different intercurrent events. Section III.D (A.3.4) gives some considerations on selecting strategies to construct an estimand.

Treatment Policy Strategy

The occurrence of the intercurrent event is considered irrelevant in defining the treatment effect of interest; the values for the variable of interest are used regardless of whether the intercurrent event occurs. For example, when specifying how to address the use of additional medication as an intercurrent event, the values of the variable of interest should be used, whether the patient takes additional medication or not.

If applied in relation to whether a patient continues treatment, and whether a patient experiences changes in other treatments (e.g. background or concomitant treatments), the intercurrent event is considered to be part of the treatments being compared. In that case, this reflects the comparison described in the ICH E9 Glossary (under ITT Principle) as the effect of a treatment policy.

In general, the treatment policy strategy cannot be implemented for intercurrent events that are terminal events, since values for the variable after the intercurrent event do not exist. For example, an estimand based on this strategy cannot be constructed with respect to a variable that cannot be measured because of death.

Hypothetical Strategies

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

A wide variety of hypothetical scenarios can be envisaged, but some scenarios are likely to be of more clinical or regulatory interest than others. For example, it may be of clinical or regulatory importance to consider the effect of a treatment under different conditions from those of the trial that can be carried out. Specifically, when additional medication must be made available for ethical reasons, a treatment effect of interest might concern the outcomes if the additional medication was not available. A very different hypothetical scenario might postulate that intercurrent events would not occur, or that different intercurrent events would occur. For example, for a subject that will suffer an adverse event and discontinue treatment, it might be considered whether the same subject would not have the adverse event or could continue treatment in spite of the adverse event. The clinical and regulatory interest of such hypotheticals is limited and would usually depend on a clear understanding of why and how the intercurrent event or its consequences would be expected to be different in clinical practice than in the clinical trial.

If a hypothetical strategy is proposed, it should be made clear what hypothetical scenario is envisaged. For example, wording such as “if the patient does not take additional medication” might lead to confusion as to whether the patient hypothetically does not take additional

medication because it is not available or because the particular patient is supposed not to require it.

Composite Variable Strategies

This strategy relates to the variable of interest (see section III.C (A.3.3)). An intercurrent event is considered to be informative about the patient's outcome and is therefore incorporated into the definition of the variable. For example, a patient who discontinues treatment because of toxicity may be considered not to have been successfully treated. If the outcome variable was already success or failure, discontinuation of treatment for toxicity would simply be considered another mode of failure. Composite variable strategies do not need to be limited to dichotomous outcomes, however. For example, in a trial measuring physical functioning, a variable might be constructed using outcomes on a continuous scale, with subjects who die being attributed a value reflecting the lack of ability to function. Composite variable strategies can be viewed as implementing the ITT principle in some cases when the original measurement of the variable might not exist or might not be meaningful, but where the intercurrent event itself meaningfully describes the patient's outcome, such as when the patient dies.

Terminal events, such as death, are perhaps the most salient examples of the use of a composite strategy. If a treatment saves lives, its effect on various measures in surviving patients may be of interest, but it would be inappropriate to say that the summary measure of interest was only the average value of some numerical measure in survivors. The outcome of interest is survival along with the numerical measures. For example, progression-free survival in oncology trials measures the treatment effect on a combination of the growth of the tumor and survival.

While-on-Treatment Strategies

For this strategy, response to treatment before the occurrence of the intercurrent event is of interest. Terminology for this strategy will depend on the intercurrent event of interest, e.g., *while alive*, when considering death as an intercurrent event.

If a variable is measured repeatedly, its values up to the time of the intercurrent event may be considered relevant for the clinical question, rather than the value at the same fixed timepoint for all subjects. The same applies to the occurrence of a binary outcome of interest up to the time of the intercurrent event. 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 while the patient is exposed to treatment.

Like the composite variable strategy, the while-on-treatment strategy can hence be thought of as having an impact on the definition of the variable, in this case, by restricting the observation time of interest to the time before the intercurrent event. Particular care is important if the occurrence of the intercurrent event differs between the treatments being compared (see section III.C (A.3.3)).

Principal Stratum Strategies

This strategy relates to the population of interest (see section III.C (A.3.3)). The target population might be taken to be the *principal stratum* (see Glossary) in which an intercurrent event would occur. Alternatively, the target population might be taken to be the principal stratum in which an intercurrent event would not occur. The clinical question of interest relates to the treatment effect only within the principal stratum. For example, there might be an interest in knowing a treatment effect on severity of infections in the principal stratum of patients becoming infected after vaccination. Alternatively, a toxicity might prevent some patients from continuing the test treatment, but there might be an interest in knowing the treatment effect among patients who are able to tolerate the test treatment.

It is important to distinguish *principal stratification* (see Glossary), which is based on potential intercurrent events (for example, subjects who would discontinue therapy if assigned to the test product), from subsetting based on actual intercurrent events (subjects who discontinue therapy on their assigned treatment). The subset of subjects who experience an intercurrent event on the test treatment will often be a different subset from those who experience the same intercurrent event on control. Treatment effects defined by comparing outcomes in these subsets confound the effects of the different treatments with the differences in outcomes possibly because of the differing characteristics of the subjects.

C. Estimand Attributes (A.3.3)

The attributes below are used to construct the estimand, defining the treatment effect of interest.

The **treatment** condition of interest and, as appropriate, the alternative treatment condition to which comparison will be made (referred to as *treatment* through the remainder of this document). These treatments might be individual interventions; might be combinations of interventions administered concurrently, e.g., as add-on to standard of care; or might consist of an overall regimen involving a complex sequence of interventions (see Treatment Policy Strategy and Hypothetical Strategies in section III.B (A.3.2)).

The **population** of patients targeted by the clinical question. This attribute might be represented by the entire trial population, a subgroup defined by a particular characteristic measured at baseline, or a principal stratum defined by the occurrence (or nonoccurrence, depending on context) of a specific intercurrent event (see Principal Stratum Strategies in section III.B (A.3.2)).

The **variable** (or endpoint) to be obtained for each patient that is used to address the clinical question. The specification of the variable might include whether the patient experiences an intercurrent event (see Composite Variable Strategies and While-on-Treatment Strategies in section III.B (A.3.2)).

Precise specifications of treatment, population, and variable are likely to address many of the intercurrent events considered in sponsor and regulator discussions of the clinical question of interest. The clinical question of interest in respect to any **other intercurrent events** will

usually be reflected using the strategies introduced as treatment policy, hypothetical or while on treatment.

Finally, a **population-level summary** for the variable should be specified, providing a basis for comparison between treatment conditions.

When defining a treatment effect of interest, it is important to ensure that the definition identifies an effect because of treatment and not because of potential confounders such as differences in duration of observation or patient characteristics.

D. Considerations for Constructing an Estimand (A.3.4)

The clinical questions of interest and associated estimands should be specified at the initial stages of planning any clinical trial. Precise specification of objectives for most trials should reflect discontinuation of treatment and use of additional or alternative treatments. In some settings, terminal events, such as death, should be addressed. Some trial objectives can only be described with reference to clinical events, for example, the duration of response in subjects who achieve a response.

The construction of an estimand should consider what is of clinical relevance for the particular treatment in the particular therapeutic setting. Considerations include the disease under study; the clinical context (e.g., the availability of alternative treatments); the administration of treatment (e.g., one-off dosing, short-term treatment, or chronic dosing); and the goal of treatment (e.g., prevention, disease modification, symptom control). Also important is whether an estimate of the treatment effect can be derived that is reliable for decision-making. For example, a clinical question on the treatment effect on clinical outcome regardless of which other therapies are used before that outcome is experienced, differs from a clinical question on the treatment effect had no additional medication been available. Depending on the setting, either treatment effect might represent a clinical question of interest. However, in both cases, a clinical trial designed to estimate these treatment effects will often include the possibility to use additional medications if medically necessary. For the former question, values after the use of additional treatment will be relevant. For the latter question, values after the additional treatment are not directly relevant since the values also reflect the impact of that additional medication. It should be agreed that reliable estimation is possible before the choice of estimand is finalized. This includes, for the latter question, the methods to replace observations that should not be used in the analysis.

When constructing the estimand, it is important to have a clear understanding of the treatment to which the clinical question of interest pertains (see section III.C (A.3.3)). Clear specifications for the treatments of interest might already reflect multiple relevant intercurrent events. Specifically, a treatment might already reflect the clinical question of interest in respect to changes in background treatment, concomitant medications, use of additional or later-line therapies, treatment-switching and conditioning regimens. For example, it is possible to specify treatment as *intervention A added to background therapy B, dosed as required*. In that case, changes to the dose of background therapy B would not need to be considered as an intercurrent event. However, the use of an additional therapy would be considered as an intercurrent event.

If the use of any additional medication is also reflected, using the treatment policy strategy for example, then treatment might be specified as *intervention A added to background therapy B, dosed as required, and with additional medication, as required*. Alternatively, if the treatment is specified as *intervention A,* then both changes in background therapy and use of additional therapy would be addressed as intercurrent events.

Discussions should also consider whether specifications for the population and variable attributes should be used to reflect the clinical question of interest in respect to any intercurrent events. Strategies can then be considered for any other intercurrent events. Usually an iterative process should be used to reach an estimand that is of clinical relevance for decision-making, and for which a reliable estimate can be made. Some estimands, in particular those for which the measurements taken are relevant to the clinical question, can often be robustly estimated, making few assumptions. Other estimands may call for methods of analysis with more specific assumptions that may be more difficult to justify and that may be more sensitive to plausible changes in those assumptions (see section V.A (A.5.1)). If significant issues exist to develop an appropriate trial design or to derive an adequately reliable estimate for a particular estimand, an alternative estimand, trial design, and method of analysis should be considered.

Avoiding or oversimplifying the process of discussing and constructing an estimand risks misalignment between trial objectives, trial design, data collection, and method of analysis. Although an inability to derive a reliable estimate might preclude certain choices of strategy, it is important to proceed sequentially from the trial objective and an understanding of the clinical question of interest, and not for the choice of data collection and method of analysis to determine the estimand.

The experimental situation should also be considered. If the management of subjects (e.g., dose adjustment for intolerance, rescue treatment for inadequate response, burden of clinical trial assessments) under a clinical trial protocol is justified to be different from that which is anticipated in clinical practice, this might be reflected in the construction of the estimand.

Once constructed, the estimand should define a target of estimation clearly and unambiguously. Consider an intercurrent event of discontinuation of treatment; it is of utmost importance to distinguish between treatment effects of interest based on the principal stratum of patients who would be able to continue if administered the test treatment and the effect during continued treatment. Furthermore, neither of these should be taken to represent an effect if all patients can continue with treatment.

As stated above, when using the hypothetical strategy, some conditions are likely to be more acceptable for regulatory decision-making than others. The hypothetical conditions described should therefore be justified for the quantification of an interpretable treatment effect that is relevant to inform the decisions to be taken by regulators, and the use of the medicine in clinical practice. 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. 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 justifiable as being of clinical or regulatory interest. A clinical question

of interest based on the effect if all subjects had been able to continue with treatment is not well-defined without a thorough discussion of the hypothetical conditions under which it is supposed that they would have continued. The inability to tolerate a treatment may itself constitute evidence of an inability to achieve a favorable outcome.

Characterizing beneficial effects using 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 upon 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 also include 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. When constructing estimands based on the treatment policy strategy, inference can be complemented by defining an additional estimand and analysis pertaining to each intercurrent event for which the strategy is used, for example, contrasting both the treatment effect on a symptom score and the proportion of subjects using additional medication under each treatment. Similarly, an estimand using a while-on-treatment strategy should usually be accompanied by the additional information on the time to intercurrent event distributions, and an estimand based on a principal stratum would usefully be accompanied by information on the proportion of patients in that stratum, if available.

The considerations informing the construction of estimand to support regulatory decision-making based on a non-inferiority or equivalence objective may differ from those for the choice of estimand for a superiority objective. As explained in ICH E9, the problem facing the regulator in its decision-making is different when based on non-inferiority or equivalence studies compared to superiority studies. In section III.B (A.3.2), it is stated that such trials are not conservative in nature and the importance of minimizing the number of protocol violations and deviations, nonadherence, and study withdrawals is indicated. In section V.B.1 (A.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 analysis of the FAS under the ITT principle. Responses in both treatment groups can appear more similar following discontinuation of randomized treatment or use of another medication for reasons that are unrelated to the similarity of the initially randomized treatments. Estimands could be constructed to directly address those intercurrent events that can lead to the attenuation of differences between treatment arms (e.g., discontinuations from treatment and use of additional medications). When selecting strategies, it might be important to distinguish between trials designed to detect whether differences exist between treatments containing the same or similar active substance (e.g., comparison of a biosimilar to a reference treatment) and trials where a non-inferiority or equivalence hypothesis is used to establish and quantify evidence of efficacy. An estimand can be constructed to target a treatment effect that prioritizes sensitivity to detect differences between treatments, if appropriate for regulatory decision-making.

IV. IMPACT ON TRIAL DESIGN AND CONDUCT (A.4)

The design of a trial should be aligned to the estimands that reflect the trial objectives. A trial design that is suitable for one estimand might not be suitable for other estimands of potential importance. 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. This includes determining the inclusion and exclusion criteria that identify the target population; the treatments, including the medications that are allowed and those that are prohibited in the protocol; and other aspects of patient management and data collection. If interest lies, for example, in understanding the treatment effect regardless of whether a particular intercurrent event occurs, a trial in which the variable is collected for all subjects is appropriate. Alternatively, if the estimands that 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 estimation, including data that inform the characterization, occurrence, and timing of intercurrent events. Data cannot always be collected. Certainly, subjects cannot be retained in a trial against their will, and in some trials, missing data for some subjects is inevitable by design, such as administrative censoring in trials with survival outcomes. On the contrary, the occurrence of intercurrent events, such as discontinuation of treatment, treatment switching, or use of additional medication, does not imply that the variable cannot be measured thereafter, though the measures may not be relevant. For terminal events such as death, the variable cannot be measured after the intercurrent event, but neither should these data generally be regarded as missing.

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 (see section V (A.5)). A prospective plan to collect informative reasons for why data intended for collection are missing may help to distinguish the occurrence of intercurrent events from missing data. This in turn may improve the analysis and may also lead to a more appropriate choice of sensitivity analysis. For example, *loss to follow-up* may more accurately be recorded as *treatment discontinuation due to lack of efficacy*. When that reason for discontinuation has been defined as an intercurrent event, this can be reflected through the strategy chosen to account for that intercurrent event and not as a missing data problem. To reduce missing data, measures can be implemented to retain subjects in the trial. However, measures to reduce or avoid intercurrent events that would normally occur in clinical practice risk reducing the external validity of the trial. 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.

Randomization and blinding remain cornerstones of controlled clinical trials. Design techniques for avoiding bias are addressed in section II.C (2.3) of ICH E9. Certain estimands may warrant, or may benefit from, use of trial designs such as run-in or enrichment designs, randomized

withdrawal designs, or titration designs. It might be of interest to identify the principal stratum of subjects who can tolerate a treatment using a run-in period, in advance of randomizing those subjects between test treatment and control. Dialogue between regulator and sponsor should 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, randomization) support 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.

A precise description of the treatment effects of interest should inform sample size calculations. Particular care should be taken when making reference to historical studies that might, implicitly or explicitly, have reported estimated treatment effects or variability based on a different estimand. When all subjects contribute information to the analysis, and when the impact of the strategy to reflect intercurrent events is included in the effect size that is targeted and the expected variance, it is not usually necessary to additionally inflate the calculated sample size by the expected proportion of subject withdrawals from the trial.

Section VII.B (7.2) of ICH E9 addresses issues related to summarizing data across clinical trials. The importance of having consistent definitions for the variables of interest is highlighted, and this can be extended to the construction of estimands. Hence, in situations when synthesizing evidence from across a clinical trial program 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 design of the contributing trials. Similar considerations apply to the design of a meta-analysis, using estimated effect sizes from completed trials to determine non-inferiority margins, 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.

More generally, a trial is likely to have multiple objectives translated into multiple estimands, each associated with statistical testing and estimation. The multiplicity issues arising should be addressed.

V. IMPACT ON TRIAL ANALYSIS (A.5)

A. Main Estimation (A.5.1)

An estimand for the effect of treatment relative to a control will be estimated by comparing the outcomes in a group of subjects on the treatment to those in a similar group of subjects on the control. For a given estimand, an aligned method of analysis, or estimator, should be implemented that is able to provide an estimate on which reliable interpretation can be based. The method of analysis will also support calculation of confidence intervals and tests for statistical significance. An important consideration for whether an interpretable estimate will be available is the extent of assumptions that are made in the analysis. Key assumptions should be stated explicitly together with the estimand and accompanying main and sensitivity estimators. Assumptions should be justifiable and implausible assumptions should be avoided. The

robustness of the results to potential departures from the underlying assumptions should be assessed through an estimand-aligned sensitivity analysis (see section V.B (A.5.2)). Estimation that relies on many or strong assumptions calls for more extensive sensitivity analysis. When the impact of deviations from assumptions cannot be comprehensively investigated through sensitivity analysis, that particular combination of estimand and method of analysis might not be acceptable for decision-making.

All methods of analysis rely on assumptions, and different methods may rely on different assumptions even when aligned to the same estimand. Nevertheless, some kinds of assumption are inherent in all methods of analysis aligned to estimands that use each of the different strategies outlined, for example, the methodology for predicting the outcomes that would have been observed in the hypothetical scenario, or for identifying a suitable target population in a principal stratum strategy. Some examples are given below related to the different strategies used to reflect the occurrence of intercurrent events. The issues highlighted will be key components of discussion between sponsor and regulator in advance of an estimand, main analysis, and sensitivity analysis being agreed upon.

Analysis aligned with a treatment policy strategy to address a given intercurrent event may entail stronger or weaker assumptions depending on the design and conduct of the trial. When most subjects are followed up even after the respective intercurrent event (e.g., discontinuation of treatment), the remaining problem of missing data may be relatively minor. In contrast, when observation is terminated after an intercurrent event, which is obviously undesirable in respect to this strategy, the assumption that (unobserved) outcomes for discontinuing subjects are similar to the (observed) outcomes for those who remain on treatment will often be implausible. An alternative approach to handle the missing data should be justified, and sensitivity analysis will be expected.

Analysis aligned to a hypothetical strategy involves outcomes different from those actually observed, for example, outcomes if rescue medication had not been given when in fact it was. Observations before the rescue medication and observations on subjects who did not need rescue medication may be informative, but only under strong assumptions.

A composite variable strategy can avoid statistical assumptions about data after an intercurrent event by considering occurrence of the intercurrent event as a component of the outcome. The potential concern relates less to assumptions for estimation, and more to the interpretation of the estimated treatment effect. For the estimand to be interpretable, if scores are assigned for failure because the intercurrent event occurs, these should meaningfully reflect the lack of benefit to the patient (e.g., death may be reflected differently than discontinuation of treatment because of adverse event).

Estimands constructed based on a while-on-treatment strategy can be estimated provided outcomes are collected up to the time of the intercurrent event. Again, the crucial assumptions concern interpretation. Consider, for example, discontinuation of treatment. Outcomes while on treatment may be improved, but the treatment may also shorten or lengthen the treatment period by provoking or delaying discontinuations. Both these effects should be considered in interpreting and assessing clinical benefit.

Analysis aligned to a principal stratum strategy usually requires strong assumptions. For example, some principal stratification methods infer this from baseline characteristics of the subjects, but the correctness of this inference may be difficult to assess. This difficulty cannot be avoided by simplified methods, however. For example, simply comparing subjects who do not have an intercurrent event on the test treatment to those who do not have an event on control, assuming intercurrent events are unrelated to treatment, is very difficult to justify.

Even after defining estimands that address intercurrent events in an appropriate manner and making efforts to collect the data needed for estimation (see section IV (A.4)), some data may still be missing, including, e.g., administrative censoring in trials with survival outcomes. Failure to collect relevant data should not be confused with the choice not to collect, or to collect and not to use, data made irrelevant by an intercurrent event. 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 therefore, for the purpose of that second estimand, are not missing if uncollected. When efforts to collect data are not successful, it becomes necessary to make assumptions to handle the missing data in the statistical analysis. Handling of missing data should be based on clinically plausible assumptions and, where possible, guided by the strategies employed in the description of the estimand. The approach taken may be based on observed covariates and post-baseline data from individual subjects and from other similar subjects. Criteria to identify similar subjects might include whether or not the intercurrent event has occurred. For example, for subjects who discontinue treatment without further data being collected, a model may use data from other subjects who discontinued treatment but for whom data collection has continued.

B. Sensitivity Analysis (A.5.2)

1. Role of Sensitivity Analysis (A.5.2.1)

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. Sensitivity analysis should be planned for the main estimators of all estimands that will be important for regulatory decision-making and labeling in the product information. This can be a topic for discussion and agreement between sponsor and regulator.

The statistical assumptions that underpin the main estimator should be documented. One or more analyses, focused on the same estimand, should then be prespecified to investigate these assumptions with the objective of verifying whether the estimate derived from the main estimator is robust to departures from its assumptions. This might be characterized as the extent of departures from assumptions that change the interpretation of the results in terms of their statistical or clinical significance (e.g., tipping point analysis).

Distinct from sensitivity analysis, when investigations are conducted with the intent of exploring robustness of departures from assumptions, other analyses that are conducted to more fully

investigate and understand the trial data can be termed *supplementary analysis* (see Glossary; section V.C (A.5.3)). When the primary estimand(s) of interest are agreed upon between sponsor and regulator, the main estimator is prespecified unambiguously, and the sensitivity analysis verifies that the estimate derived is reliable for interpretation, supplementary analyses should generally be given lower priority in assessment.

2. Choice of Sensitivity Analysis (A.5.2.2)

When planning and conducting a sensitivity analysis, altering multiple aspects of the main analysis simultaneously can make it challenging to identify which assumptions, if any, are responsible for any potential differences seen. It is therefore desirable to adopt a structured approach, specifying the changes in assumptions that underlie the alternative analyses, rather than simply comparing the results of different analyses based on different sets of assumptions. The need for analyses varying multiple assumptions simultaneously should then be considered on a case-by-case basis. A distinction between testable and untestable assumptions may be useful when assessing the interpretation and relevance of different analyses.

The importance of sensitivity analysis in respect to missing data is established and retains its importance in this framework. Missing data should be defined and considered in respect to a particular estimand (see section IV (A.4)). The distinction between data that are missing in respect to a specific estimand and data that are not directly relevant to a specific estimand gives rise to separate sets of assumptions to be examined in sensitivity analysis.

C. Supplementary Analysis (A.5.3)

Interpretation of trial results should focus on the main estimator for each agreed-upon estimand, providing that the corresponding estimate is verified to be robust through the sensitivity analysis. Supplementary analyses for an estimand can be conducted in addition to the main and sensitivity analysis to provide additional insights into the understanding of the treatment effect. They generally play a lesser role for interpretation of trial results. The need for, and utility of, supplementary analyses should be considered for each trial.

Section V.B.3 (5.2.3) of ICH E9 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. It is also described in Section V.B.2. (5.2.2) that results based on a PPS might be subject to severe bias. In respect to the framework presented in this addendum, it may not be possible to construct a relevant estimand to which analysis of the PPS is aligned. As noted above, analysis of the PPS does not achieve the goal of estimating the effect in any principal stratum, for example, in those subjects able to tolerate and continue to take the test treatment, because it may not compare similar subjects on different treatments.

Protocol violations and deviations might exclude subjects from the PPS, for example, by having a visit outside a time window, without an intercurrent event necessarily having occurred. Likewise, subjects could experience an intercurrent event, such as death, without having deviated

from the protocol. Notwithstanding the differences between violations and deviations from the protocol and intercurrent events, events likely to affect the interpretation or existence of measurements are considered in the description of the estimand. Estimands might be constructed with aligned methods of analysis that better address the objective usually associated with the analysis of the PPS. If so, analysis of the PPS might not add additional insights.

VI. DOCUMENTING ESTIMANDS AND SENSITIVITY ANALYSIS (A.6)

A trial protocol should define and specify explicitly a primary estimand that corresponds to the primary trial objective. The protocol and the analysis plan should prespecify 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 also be defined and specified explicitly, each with a corresponding main estimator and a suitable sensitivity analysis. Additional exploratory trial objectives may be considered for exploratory purposes, leading to additional estimands.

The choice of the primary estimand will usually be the main determinant for aspects of trial design, conduct, and analysis. Following usual practices, these aspects should be well documented in the trial protocol. If secondary 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.

Although it is to the benefit of the sponsor to have clarity on what is being estimated, it is not a regulatory requirement to document an estimand for each exploratory objective.

Results from the main, sensitivity, and supplementary analyses should be reported systematically in the clinical trial report, specifying whether each analysis was prespecified, introduced while the trial was still blinded, or performed post hoc. Summaries of the number and timings of each intercurrent event in each treatment group should be reported.

Changes to the estimand during the trial can be problematic and can reduce the credibility of the trial. Addressing intercurrent events that were not foreseen at the design stage, and are identified during the conduct of the trial, should discuss not only the choices made for the analysis, but the effect on the estimand (i.e., on the description of the treatment effect that is being estimated) and the interpretation of the trial results. A change to the estimand should usually be reflected through amendment to the protocol.

GLOSSARY

Estimand:
A precise description of the treatment effect reflecting the clinical question posed by the trial objective. It summarizes at a population-level what the outcomes would be in the same patients under different treatment conditions being compared.

Estimate:
A numerical value computed by an estimator.

Estimator:
A method of analysis to compute an estimate of the estimand using clinical trial data.

Intercurrent Events:
Events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest. Intercurrent events should be addressed when describing the clinical question of interest to precisely define the treatment effect that is to be estimated.

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:
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:
In this document, a *principal stratum* refers to any of the strata (or combination of strata) defined by principal stratification.

Sensitivity Analysis:
A series of analyses conducted with the intent to explore the robustness of inferences from the main estimator to deviations from its underlying modeling assumptions and limitations in the data.

Supplementary Analysis:
A general description for analyses that are conducted in addition to the main and sensitivity analysis with the intent to provide additional insights into the understanding of the treatment effect.