

**Final Concept Paper**  
**E9(R1): Addendum to Statistical Principles for Clinical Trials**  
**on**  
**Choosing Appropriate Estimands and Defining Sensitivity Analyses in Clinical Trials**  
**dated 22 October 2014**  
*Endorsed by the ICH Steering Committee on 23 October 2014*

### **Type of Harmonisation Action Proposed**

To develop new regulatory guidance, suggested to be an Addendum to ICH E9, which promotes harmonised standards on the choice of estimand in clinical trials and describes an agreed framework for planning, conducting and interpreting sensitivity analyses of clinical trial data. As with ICH E9, the Addendum will focus on statistical principles related to estimands and sensitivity analysis, not on the use or acceptability of specific statistical procedures or methods. While a variety of mid-stage and late-stage clinical trials may be in scope, the primary focus of the Addendum will be on confirmatory clinical trials.

### **Statement of the Perceived Problem**

- *Incorrect choice of estimand and unclear definitions for estimands lead to problems in relation to trial design, conduct and analysis and introduce potential for inconsistencies in inference and decision making.*

Inferences about the true efficacy and safety profile of a medicinal product are drawn from estimated effects in confirmatory clinical trials. A clinical trial protocol and analysis plan should include a ‘golden thread’ linking clear trial objectives with selection and prioritisation of endpoints and hypotheses for statistical testing or targets for estimation. These should, in turn, inform details of the trial design, conduct and analysis. In a confirmatory clinical trial data are collected to measure outcomes that quantify the impact of one or more experimental interventions in comparison to a control group, typically over a defined period of time, or until a sufficient number of clinical outcome events have occurred. The trialist is trying to formulate an appropriate and well-defined measure of treatment effect in terms of the data that were intended to be collected. This may then be parameterised, for example to “compare experimental drug X and placebo in terms of improving endpoint Y at time Z for all randomised patients, without regarding adherence to randomised treatment” or to “compare experimental drug X and placebo in terms of improving endpoint Y at time Z for all randomised patients if all patients had remained in the trial and received treatment as planned without rescue medication until time Z”. Controversy and confusion exist on the definition and appropriate selection of an appropriate estimand and these two examples should not be taken as preferences or recommendations. These are presented only as illustrations of *estimands*; the property that is to be estimated in the context of a scientific question of interest, to stimulate discussion in generating the addendum.

A clear definition of an estimand is important not only so that the analysis can be pre-specified in all main aspects, but also since the choice of estimand is linked to important considerations around trial design, conduct and analysis. These include, for example, duration of patient follow-up, adherence to randomised treatment, use of alternative

medications after discontinuation of randomised treatment and methods to handle missing data in the statistical analysis.

Remarkably, despite many years of clinical trials being the primary support for regulatory decision making, no definitive guidance is available on what constitutes an appropriate primary estimand for a confirmatory clinical trial. The absence of regulatory guidance leads to uncertainties and inconsistencies in methodological approach across trial designs supporting regulatory decisions. Furthermore, similar submissions may lead to different inferences being drawn by different regulatory authorities for reasons that appear to have nothing to do either with actual regional differences or even with clear differences of opinion in benefit-risk appraisal, but rather with a lack of what could be a common understanding of trial objectives and of what constitutes an appropriate quantification of the effects of an experimental treatment.

- ***Absence of a framework for planning, conducting and interpreting sensitivity analyses may lead to inconsistencies in inference and decision making within and between regulatory regions.***

Following ICH E9, it has become standard in all regions to pre-specify a primary statistical analysis for efficacy, but it has also been common practice to investigate the extent to which the outcomes of other approaches to the analysis lead to consistent findings. In defining an appropriate ‘estimand’ for each primary and secondary endpoint, and in determining a strategy for statistical analysis to derive estimated effects, a number of choices and assumptions need to be made. A targeted range of thoughtfully constructed ‘sensitivity analyses’ can help to investigate and understand the robustness of estimates; the sensitivity of the overall conclusions to various limitations of the data, assumptions, and approaches to data analysis. At present, while sensitivity analyses are presented, they are rarely based on a systematic consideration of the various choices and assumptions made and are rarely discussed in terms of their relative importance for decision making. There are no clear regulatory standards to follow in defining an appropriate set of sensitivity analyses or for the joint interpretation of these “sensitivity analyses” with the primary analysis. Consequently, decisions of regional authorities may appear to weight evidence in an arbitrary and unpredictable manner, even if underpinned by an agreed set of statistical principles. Greater clarity in this regard should assist applicants in planning more appropriate submissions and may increase the predictability of regulatory appraisals.

### **Issues to be Resolved**

- ***Incorrect choice of estimands and unclear definitions for estimands lead to problems in relation to trial design, conduct and analysis and introduce potential for inconsistencies in inference and decision making.***

In respect of estimands, the factors which may be used to define and describe different estimands, and the different levels of each factor, should be identified. These factors will include the outcome measure, treatment received, analysis population, time period of interest and treatment adherence status. For example, in terms of analysis population one may be interested in the full analysis set, the per-protocol set or the ‘treatment-adherers’ set; in terms of timepoint one might be interested in response at a fixed timepoint without regard to treatment adherence, or response only whilst receiving randomised treatment. In terms of treatment adherence status estimands can, in principle, be constructed to investigate the effect regardless of adherence to treatment, or the effect if treatment had been taken as planned by

the whole analysis population. From this a series of relevant estimands may be identified and harmonised guidance given on circumstances where it is appropriate to choose each one as an estimand of primary interest.

This problem has received some attention in published literature. For example, a report from the U.S. National Academy of Sciences research arm, the National Research Council (NRC) describes that important considerations for trial design are related to different ways in which trial objectives and hypotheses of potential interest might be framed, using 5 constructs for illustration:

1. (Difference in) mean outcome improvement for all randomised participants;
2. (Difference in) outcome improvement in those who adhere to treatment;
3. (Difference in) outcome improvement if all participants had adhered;
4. (Difference in) areas under the outcome curve during adherence to treatment;
5. (Difference in) outcome improvement during adherence to treatment;

*Mallinckrodt et al*, give a further illustration:

6. For all randomised participants at the planned endpoint of the trial attributable to the initially randomized treatment.

Whilst some of the constructs described will not produce a suitable estimand for use in clinical trials in support of regulatory submissions, each of these may be thought of as an illustration of a different estimand. For each estimand, one or more statistical analyses will then need to be selected, each of which will produce a different estimator. The objective of the guidance is to describe a framework in which the choice of appropriate estimand can be made and agreed between sponsor and regulator based on clear descriptions. It is anticipated that more detailed and specific definitions will be generated through discussion.

Current practice by drug developers is varied and without clear structure and it is argued that none of the estimands currently discussed are suitable across the full range of experimental situations faced by drug developers targeting regulatory submission. Considerations relating to trial design differ according to therapeutic area, and it may be expected that the estimand of primary relevance will also differ according to the experimental situation. The main concern in relation to current practice is the absence of a clear relationship between the apparent target estimand and the trial design and analysis in terms of aspects such as patient follow-up after discontinuation of randomised treatment, decisions around which data to exclude from the statistical analysis and handling of missing data..

- ***Absence of a framework for planning, conducting and interpreting sensitivity analyses may lead to inconsistencies in inference and decision making within and between regulatory regions.***

A common understanding of what is meant by sensitivity analyses should be derived and this should inform a statement to clarify the objectives for sensitivity analyses. One type of supportive analysis might investigate treatment effects according to a range of different estimands. At the present time such analyses can be useful, primarily because there is no harmonisation on the definition of an estimand of primary importance. The extent to which this type of analysis is needed once an estimand of primary importance is agreed should be discussed while developing this guidance. A second type is analyses that address departures from the choices and assumptions made in support of the primary analysis of a particular variable. For example, the NRC report (2010) lists (a) distributional assumptions for the full data, (b) outlying or influential observations, and (c) assumptions about the missing data

mechanism, as different types of sensitivity analyses. At the present time it is commonplace for sponsors to present results without precise description of the estimand of interest and to present analyses purported to be sensitivity analyses without indication of which choice or assumption is being investigated. As a minimum, the list of all planned sensitivity analyses should be presented clearly, with a specific rationale for each. Indeed, analyses described as ‘sensitivity analyses’ may in fact be based on the same underlying assumptions as the primary analysis, or may address a different estimand when purporting to examine a choice or assumption made for the primary statistical analysis. It may be discussed whether it is required that sensitivity analyses make different mathematical assumptions from the primary analysis, or whether it would suffice that they use different methods. In addition, a tension exists between providing analyses based on assumptions that may be considered more plausible, or may be considered more convenient in terms of statistical analysis, or that may be more likely to provide conservative analysis in terms of avoiding artificial inflation of the estimated effect in favour of the experimental treatment.

Inferences from clinical trial data may be complicated by the proliferation of estimated effects and p-values that are generated from multiple sensitivity analyses and these have the potential to confuse and distract decision makers. The ability to discriminate between additional analyses that constitute a thorough and accurate statistical review and those that serve to confuse remains a challenge that needs to be addressed. An improved framework should focus the sensitivity analyses that are provided to those that add value to decision making by targeting a particular assumption that is critical for inference. In addition, interpretation of sensitivity analyses should be discussed. It should be discussed whether, in order to confirm robustness, results should be consistent in terms of the strength of evidence presented from a statistical perspective and/or in terms of the size of estimated effects from a clinical perspective. It may be questioned whether it suffices for decision makers to be presented with explanations for discrepancies between different analyses.

The problem of deriving a framework for sensitivity analyses is particularly acute as concerns handling missing data. This is because assumptions commonly made in relation to the problem of missing data are, arguably, less likely to be valid, and deviations from assumptions can be of greater consequence for estimation and statistical testing. It is increasingly recognised that methods for primary analysis are proposed which focus on a particular type of ‘estimand’ and that rely on assumptions that cannot be verified and may appear implausible. There is recognition from clinical trial sponsors that these primary analyses must be supported by methods varying the assumptions in respect of missing data handling, but it is not always clear how to do this, and it is rarely done such that all important assumptions are highlighted and examined. Indeed, alternative analyses presented that purport to show robustness of findings may in fact rely on the same assumptions as for the primary analysis. Added to this, certain estimands require collection of data after discontinuation of randomised treatment, and this is not routinely practised. Nevertheless, it is proposed that the discussion of a framework for sensitivity analyses should cover these topics comprehensively and not only in relation to the problem of missing data.

## Background to the Proposal

ICH E8 gives a high level overview of different phases of clinical development, touching briefly on objectives for trials in each phase. ICH E9 describes important statistical principles for clinical trials, discussing a number of factors individually that are important in determining an appropriate estimand, including analysis sets and dealing with missing values. Reporting of results from sensitivity analysis is relevant for ICH E3.

This document also presents the idea of using sensitivity analyses to explore robustness of results, for example in the context of determining a sample size and in dealing with missing values. Neither ICH document addresses the issues described above. The same can be said of the various regional guidance documents that have been developed, though particular reference is made to the CHMP Guideline on Missing Data in Confirmatory Clinical Trials (EMA/CPMP/EWP/1776/99 Rev. 1) which expands on some of the important aspects and principles important to the issues described above and to the National Academy of Sciences Report on The Prevention and Treatment of Missing Data in Clinical Trials, which describes a series of different estimands for consideration.

These are, at least in the most part, subjects where the regional authorities do not currently have clear standards, so that it is a matter for development rather than harmonisation of standards. Of course, that makes it all the more feasible since there is no definitive regional regulatory guidance with which a harmonised guideline may be seen to conflict. It is anticipated that a harmonised guidance on these aspects would present a more transparent approach than is current in any region. It has the potential to enable applicants to better plan global submissions and to understand the decision making by different regulatory authorities.

## References

- ICH E8, General Considerations for Clinical Trials, July 1997;
- ICH E9, Statistical Principles for Clinical Trials, February 1998;
- Committee for Medicinal Products for Human Use (CHMP) Guideline on Missing Data in Confirmatory Clinical Trials (EMA/CPMP/EWP/1776/99 Rev. 1);
- National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. Washington, DC: National Academies Press, 2010;
- Lisa M. LaVange. The Role of Statistics in Regulatory Decision Making. *Therapeutic Innovation & Regulatory Science* 2014;48(1):10-9;
- Mallinckrodt CH, Lin Q, Lipkovich I, Molenberghs G.. A structured approach to choosing estimands and estimators in longitudinal clinical trials. *Pharm Stat* 2012;11(6):456-61;
- Little RJ, Cohen ML, Dickersin K, et al. The design and conduct of clinical trials to limit missing data. *Stat Med* 2012;31(28):3433-43;

Little RJ, D'Agostina R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med* 2012;367(14):1355-60.

## Type of Expert Working Group and Resources

The EWG will be comprised of two members (statisticians) nominated by EU, EFPIA, FDA, PhRMA, MHLW, JPMA, Health Canada and Swissmedic. One member can also be nominated by WHO Observer, biotech industry, as well as RHIs, DRAs/DoH (if requested). Access to clinical expertise may be required as the guidance is developed.

**Timing**

The group will start working remotely in the 3rd quarter of 2014. A *Step 1* document will be drawn up during the first EWG meeting which is anticipated at the EU meeting in November 2014. The EWG will prepare a *Step 2* document in the June 2015 meeting in Japan. The *Step 2b* document will be published for consultation in June 2015. After collecting and incorporating public comments, a *Step 4* document will be finalised in November 2015.

Adoption of Concept Paper by the ICH Steering Committee	October 22, 2014
EWG starting work remotely	3Q 2014
Establishment of a <i>Step 1</i> document	November 2014
Preparation of a <i>Step 2a/b</i> document	June 2015
Publication of the <i>Step 2b</i> document for consultation	June 2015
Finalisation of a <i>Step 4</i> document	November 2015