March 9th, 2018

Public Comments to ICH E9（R1） Estimands and Sensitivity Analysis in Clinical Trials

T. Shun Sato
Department of Biostatistics
Kyoto University School of Public Health

1. General comment

While the original ICH E9 document had 31 pages, it seems too much that its addendum has 22 pages. Please specify “statistical principles” in this addendum, delete its redundant parts, and make the addendum compact.

It seems that a new framework based on estimands is a good statistical principle. However, once a sponsor defines an estimand, how to handle intercurrent events in order to estimate it properly is a sponsor’s responsibility, and it’s out of a statistical principle.

In “7. A Generic Example”, it reads “(i)t should not be construed as a regulatory recommendation (lines 666-667),” and I do not think that the Section 7 is a part of International Council for Harmonisation Guideline. This part will be published as an external document, such as case studies.

Although its title is “Estimands and Sensitivity Analysis in Clinical Trials”, “5.2. Sensitivity Analysis” only has one page long. It seems better to shorten the descriptions of estimands and add more about sensitivity analysis.

2. “1. Purpose and Scope,” lines 45-52

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 term “the effect of a treatment policy” was only appeared in its Glossary and not in the main text of ICH E9. E9 was focused on the estimation of treatment effect based on the Full Analysis Set (ITT). It is because, in the presence of intercurrent events, there was no consensus

among the clinical trial community on the unbiased estimation of the treatment effects (average causal effects) at the time of ICH E9. (At that time, instrumental variables methods were introduced to clinical trial area.)
Thus, the E9 expert working group (EWG) adopted the ITT principle, which gives valid alpha level tests in many intercurrent event settings. ICH E9 did not recommend the estimation of the effect of a treatment policy. Instead, although it introduces a bias in the estimation of the treatment effect, the E9 EWG would have liked to adopt a statistical approach which usually keeps alpha levels of the test of no treatment effects.
(At the 2017 Joint Statistical Meeting, I talked with Professor Susan Ellenberg, who was a member of the E9 EWG on this matter, and Professor Ellenberg agreed with me.)

3. "3.2. Strategies for Addressing Intercurrent Events"

In Section 3.2, it seems that those strategies define estimands. However, alternative interpretation of these strategies is that they give biased estimators of (other) estimands. As noted in my comment 2, it is interpreted that treatment policy strategy gives a biased estimator of an estimand based on hypothetical strategy.
However, there is no such a view that a specific strategy gives a biased estimator of the other estimand (for example, an estimator based on treatment policy strategy is interpreted as a biased estimator of an estimand based on hypothetical strategy). In the addendum, an estimand, target of estimation, and an asymptotic expectation of such a biased estimator seem to be mixed up. Please describe these two differently.

4. "3.2. Strategies for Addressing Intercurrent Events", lines 186-188

Together with the other estimand attributes, the choices made on how to address intercurrent events describe the treatment effect that is targeted.

In the addendum, it is clearly noted that define an estimand first, and that align design, conduct, and analysis as summarized in Figure 1. However, there are many descriptions that strategies addressing intercurrent events define estimands in the addendum. The statistical principle is that define an estimand first, and that choose a proper strategy for addressing intercurrent events in order to estimate the defined estimand unbiasedly. Descriptions in this addendum should follow this statistical principle as summarized in Figure 1.
Sometimes, we may not or be hard to have any proper strategy for addressing intercurrent events to estimate the defined estimand unbiasedly. In such cases, as is noted in my comment 3, we may choose a strategy to give a biased estimator of the defined estimand, instead of choosing an alternative estimand.

5. “3.2. Strategies for Addressing Intercurrent Events”, Composite strategy

I do not think that composite strategy is not a part of statistical principles in this addendum. Composite strategy is not a strategy for addressing intercurrent events themselves, but a strategy to alter intercurrent events under the defined estimand to a part of a new primary variable with changing the definition of the primary variable. Changing the definition of the primary variable means changing the estimand. Such a strategy clearly be against the statistical principle in Figure 1. Delete “composite strategy,” or explain it from the view of providing a biased estimator of the defined estimand.

6. “3.2. Strategies for Addressing Intercurrent Events”, lines 227-234

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.

When death is “much more meaningful than observations before death,” death itself should be a part of the primary variable, not an intercurrent event. This paragraph only mentioned on composite endpoints.

7. “3.3.1. General Considerations”, lines 310-312

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.

A multi-disciplinary nature of the construction of the estimand is a very important message in the addendum. However, when the description of estimands is only given in the addendum, non-statistician clinical trial staff will not notice this important message. The notion of estimands should be given in ICH E6 and E8, otherwise other clinical trial staff may not know the importance of estimands. Please propose this from Japan in the following process.

8. “3.3.1. General Considerations”, lines 327-330

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.

Considering an alternative estimand means that a primary objective of a planned trial cannot be achieved. And hence, a trial using “an alternative estimand, trial design and analytic approach” may not be considered as an important trial for licensing by regulatory agencies.
As noted in my comment 2, a biased estimator of “a particular estimand” is much more meaningful than “an alternative estimand.” The addendum lacks such a point of view.

9. “3.3.2. Considerations of Therapeutic and Experimental Context”, lines 373-374

This dichotomisation of continuous scores would thus result in a change of the estimand.

In the addendum, descriptions such that strategies for addressing intercurrent events define estimands seem everywhere. Considering the statistical principles in the addendum, I do not understand the allowable situation of changing the defined estimand by dichotomization of continuous scores. This point clearly shows a problem of composite strategy in my comment 5. Justification of dichotomization should be that the trial objective is to estimate such a particular estimand (although I do not know what kind), and hence it is not the composite strategy for addressing intercurrent events.

10. “4. Impact on Trial Design and Conduct”, lines 469-473

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.

Would it be possible not to choose randomization in a confirmatory trial?

11 “4. Impact on Trial Design and Conduct”, lines 495-500

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.

Does it mean that adjustment for multiplicity is required for multiple estimands? I do not understand how to adjust multiplicity of estimands. When “a trial design that is suitable for one estimand might not be suitable for other estimands of potential importance,” it’s not only a problem for multiplicity, but also a problem of bias.

12. “5.3. Supplementary Analysis”, lines 624-625

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.

One of analytical approach to handle intercurrent events is an instrumental variables method. In randomized clinical trials, randomization is considered as an instrumental variable and one can calculate bounds for the treatment effect. (see, Robins and Greenland, JASA 1996; 91: 456-458, Greenland, International Journal of Epidemiology 2000; 29: 722-729.)
Further assumptions may narrow bounds between the treatment effect under the FAS (ITT) analysis and that under the PPS analysis (Chiba, Sato, and Greenland, Statistics in Medicine 2007; 26: 5125-5135). Since we can consider the FAS analysis and the PPS analysis as a part of sensitivity analysis, this sentence seems too strong.