EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

13 December 2018
EMA/845963/2018
Human Medicines Research and Development Support Division

Question and answer on adjustment for cross-over in estimating effects in oncology trials

Agreed by Biostatistics Working Party	November 2018
Adoption by CHMP	13 December 2018

Keywords	***Cross-over, treatment switch, oncology, estimand, target of estimation, bias, intention-to-treat (ITT), assumption, hypothetical, adjustment, censoring, inverse probability of censoring weighting (IPCW), rank preserving structural failure time (RPSFT), two-stage***

30 Churchill Place ● Canary Wharf ● London E14 5EU ● United Kingdom
Telephone +44 (0)20 3660 6000 **Facsimile** +44 (0)20 3660 5510
Send a question via our website www.ema.europa.eu/contact

Question

Which methods could be used for the analysis of time-to-event endpoints when cross-over occurs in an oncology trial, where the most prominent example is that part of the control patients cross over to the experimental treatment when progression is observed and the interest is in estimating the hypothetical effect on OS when none of the control patients would have crossed over.

Additionally questions are:

- What are the (major) available methods for dealing with cross-over and what scientific questions do they try to answer?
    - Are there underlying assumptions as regards patients crossing over and those not crossing over?
    - Are there assumptions regarding the activity of the 'crossed-over-to' treatment?
    - Are there other important assumptions of relevance for the interpretation of the results?
- Which methods and in which situations are (more) appropriate (than other methods)?
- Which role can these analyses play in regulatory decision-making?

Answer

In oncology trials, one-sided cross-over of control patients to the experimental treatment may occur, e.g. after progression. No objections from a methodological perspective exist against systematic cross-over, where systematic means that there is an objective criterion which determines whether a control patient will cross over to the experimental treatment, or not. One example is when all control patients switch to experimental treatment at the same calendar time (e.g. after an interim analysis declaring superiority); however, this is provided that unconfounded overall survival (OS) data are not considered necessary to evaluate efficacy or safety. Another example of systematic cross-over is when by design a control patient must switch to experimental treatment when that patient experiences progression and the outcome is another measure than progression, e.g. OS, if it is justified to use this design. Non-systematic cross-over can occur, for instance, when the study protocol allows cross-over after progression *at the discretion of the investigator.* This document addresses the situation where cross-over of control patients is not systematic and there is interest in estimating the effect in the (hypothetical) situation that no cross-over would have occurred in the trial, under the assumption that the experimental treatment cannot introduce harm or deterioration of the condition under investigation in the control patients who cross over. In particular, it should be fully justified that this hypothetical effect is a relevant one for regulatory decision making. It should be noted that due to the uncertainties involved in the methods described below, such estimations should, at present, be used primarily as supportive or sensitivity analyses.

Deciding on the question of interest

Different statistical methods have been proposed to adjust overall survival for cross-over, including analysis censoring at time of cross-over, Inverse Probability of Censoring Weighting (IPCW), Rank Preserving Structural Failure Time models (RPSFT), and 'two-stage' methods.

In principle, these methods can (be adapted to) address different questions by formulating distinct estimands. For example, IPCW estimates the effect of the experimental treatment versus control as if

cross-over by control group patients to the experimental treatment was absent but still includes subsequent therapies. Using RPSFT the analyst could choose the estimate to aim at the effect of experimental therapy only (effect of being 'on experimental treatment'), but in practice the effect of experimental therapy and subsequent therapies (effect of 'ever being treated') is often estimated.

Therefore, it should be clear which question is relevant and how the chosen analysis answers it.

Assumptions of adjustment methods

To arrive at effect estimates, adjustment methods have to make assumptions. To elucidate these, a short description of the working of the methods and their assumptions follows.

Censoring at cross-over

Censoring a patient at time of cross-over effectively estimates the survival of patients in absence of cross-over as that of other patients that have not crossed over at that time point and are in the same treatment arm. Therefore, the underlying assumption is that the control patient crossing over at time t has the same OS prognosis from time t onwards as the patients remaining on control treatment at time t ('non-informative censoring'). This may be a reasonable assumption, for example if all control patients cross over after an interim analysis. In general, this is an unrealistic assumption, as cross-over is often related to the patient prognosis. For example, when the prognostically better patients cross over, censoring at cross-over assumes that their survival would be as that of the remaining, prognostically worse patients, so it underestimates their survival times. This leads to an over-estimate of the OS treatment effect for the hypothetical situation that none of the control patients would have crossed over. In fact, censoring is often anti-conservative and cannot be recommended. Below several proposals are discussed to better account for informative censoring, although they still need specific assumptions.

Inverse Probability of Censoring Weighting (IPCW)

Each time a control patient crosses over to the experimental treatment, IPCW censors that patient and let the remaining control patients that are similar in terms of specified (baseline and post-baseline) characteristics, count for more patients to replace that patient. Technically, the remaining control patients are upweighted to effectively construct a pseudo-population that has the same specified (baseline and post-baseline) characteristics as the original population, but did not cross over. This will only provide unbiased estimates if the created pseudo-population has the same prognosis as the original population, and this is an untestable assumption ('no unmeasured confounders' assumption). Furthermore, sufficient heterogeneity in the cross-over pattern should be present in the sense that baseline and time-dependent covariates should not be (close to) perfectly predicting cross-over; for if they do, the weights used in IPCW would be (close to) infinity, and the method would break down.

Rank Preserving Structural Failure Time models (RPSFT)

RPSFT models assume that each patient proceeds through the disease process towards his/her death at his/her own speed (accelerated failure time model) and that the experimental therapy slows this speed down by the same factor regardless whether experimental therapy is given from randomization or from the time of cross-over ('common treatment effect'). Then RPSFT applies this slow-down factor to cross-over patients to recover which event times would have been observed had these patients not crossed over. From this, the effect in absence of cross-over can then be estimated.

The assumption that this slow-down factor is not dependent on time of starting experimental therapy ('common treatment effect' assumption) may be questionable. Moreover, RPSFT models split the

survival time in time 'on' and 'off' the experimental treatment in order to estimate the slow-down factor. Choosing subsequent therapies as 'off' treatment assumes that these are similarly effective as the control, while considering therapies subsequent to the experimental treatment as still 'on' treatment estimates an average effect of experimental and subsequent therapies.

Two-stage methods

'Two-stage' methods first estimate the effect specific to patients crossing over (by comparing control patients who do and do not cross-over) and use this effect for adjusting survival times for patients that did cross over to obtain their counterfactual survival times in absence in cross-over. A simplified procedure to do this introduces a 'secondary' baseline, often the time of progression. The effect of treatment in all control patients from progression onwards (i.e. after crossing over) is estimated using an accelerated failure time model on the post-progression data in the control group. Differences in covariates measured at time of progression may be adjusted for. The resulting slow-down factor is inversely applied to control patients that switched over to obtain counterfactual event times for those patients in absence of cross-over. In contrast to RPSFT the common treatment effect assumption is not required. More complex 'two-stage' models would also try to adjust for differences that arise over time between secondary baseline and time of actual cross-over, not only those at secondary baseline. Assumptions needed are therefore the 'no unmeasured confounders' assumption at secondary baseline for the simplified model and in addition that there are 'no unmeasured time-dependent confounders' for the more complex models.

Discussion

While in general these adjustment methods will tend to have less bias for the 'ITT in absence of cross-over' effect than the '(pure) ITT' analysis (i.e., analysis of OS irrespective of cross-over), they are often not more precise. For example, RPSFT models will typically not change the p-value, and while IPCW and 'two-stage' methods might, confidence intervals for all three methods tend to be wide.

Which method will be more appropriate than others depends on which assumption seems most reasonable or robust to deviations. In general, it would be challenging to make reasonable assumptions on the activity of the experimental treatment after cross-over based on data from the control arm which is a limitation of the RPSFT adjustment. As for IPCW and 'two-stage' methods, numerous 'subjective' factors influence which patients are selected by doctors to cross over. Therefore, it will in practice hardly be possible to convincingly justify the 'no unmeasured confounders assumptions' as it requires the set of factors determining cross-over to be known and measured at appropriate time points in the trial.

Conclusion

In principle, the scientific question of interest guides which analyses are of interest, and it should be fully justified that the chosen question is the relevant one for regulatory decision making. Justification for targeting the survival outcome if no cross-over had occurred should include discussing why preference is given to this approach over the assessment of survival at the end of the trial irrespective (or with a specific discussion about the importance) of cross-over. If it can be justified that the regulatory/clinical question of interest is what would have been the outcome in terms of survival without cross-over, a careful discussion is needed. This should address: a) whether the pivotal underlying assumptions of the chosen analysis method(s) can be regarded as plausible; b) whether the potential bias of these methods under these assumptions is sufficiently small with regard to the context at hand; c) whether these methods are reasonably robust in case of realistic deviations from these assumptions; d) supporting ancillary evidence (e.g. understanding of the mechanism of action).

Given that the underlying assumptions of the adjustment methods for cross-over described above can in principle not be proven to be true, a positive result from an analysis adjusted for cross-over cannot be used to rescue a trial that is negative as per other evidence, or to ascertain that a treatment confers an OS advantage when this is not apparent in an analysis that does not (strongly) depend on unverifiable assumptions, such as an 'ITT-analysis' that uses the observed OS outcome for each patient. For these reasons, these analyses may only be useful for regulatory purposes as supportive or sensitivity analyses with (as outlined above) a clearly demonstrated robustness against deviations from the underlying assumptions.

References:

1. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. ICH Harmonised Guideline. Current Step 2 version dated 16 June 2017.
2. Latimer NR, Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. 2014.
3. Fojo T, Amiri-Kordestani L, Bates SE. Potential Pitfalls of Crossover and Thoughts on Iniparib in Triple-Negative Breast Cancer. J Natl Cancer Inst 2011;103:1738–1740