PhUSE 2015

Paper SP02

Getting closer to the truth? Adjusting long-term efficacy estimates when patients cross over to experimental treatment

Sabine Lauer, Dr. Lauer Research, Neu-Isenburg, Germany

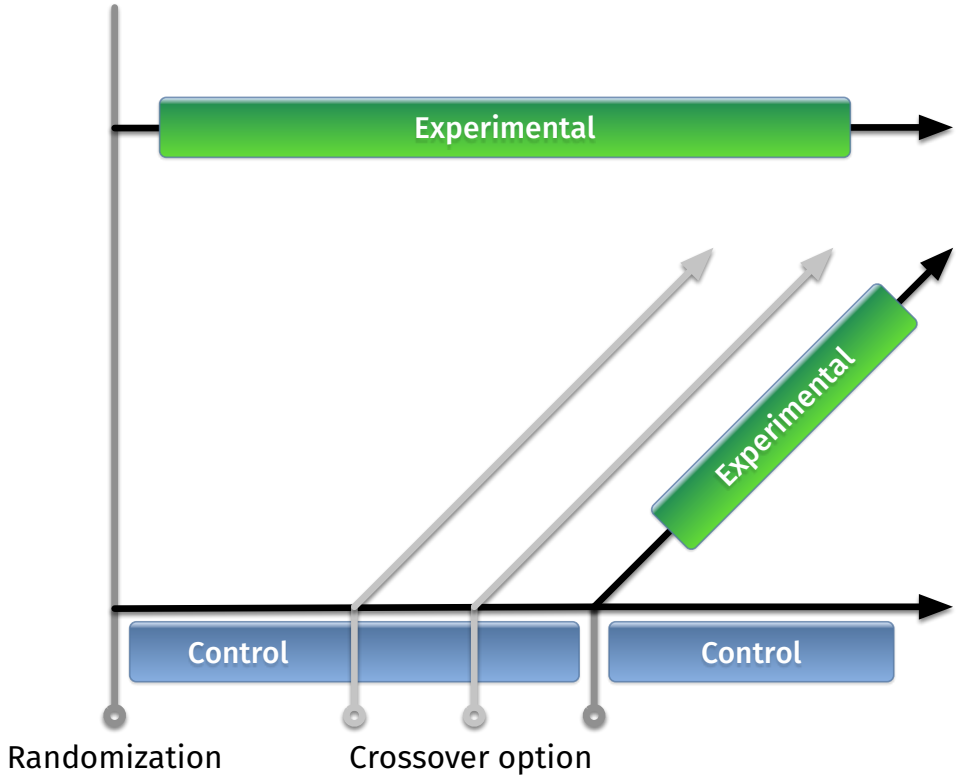
ABSTRACT

In clinical rials, patients randomized to the control arm are sometimes allowed to cross over to experimental treatment e.g. after successful interim analysis or following disease progression. Using the conventional ITT approach likely underestimates the true treatment effect and may have serious impact on the power of the trial. We present several methods trying to adjust for crossover in the sense of estimating the treatment effect that would have been observed if no crossover had occurred. Focus is given to the rank-preserving structural failure model discussing the statistical and practical aspects of the use of this method.

We conclude that the treatment effect may be strongly underestimated if no corrective steps are taken. However, we also observe that there is currently no method that could be seen as "gold standard". The specifics of each trial need to be taken into account for selecting a method and applying/fine-tuning the selected method.

INTRODUCTION

Crossover is a common phenomenon in oncology trials – with patients in the control arm switching to the new therapy during the course of the trial.



The switching option is generally introduced after a strong efficacy signal in an early interim analysis. This surely confounds the usual intention-to-treat (ITT) analysis when patients are analyzed to their randomized treatment and leads to an underestimate of the true treatment effect in later analyses. Ideally, we should best not allow at all for crossover but it is sometimes hard to avoid to due ethical reasons. Can we adjust for crossover via suitable statistical methods to get a better understanding of the magnitude of treatment benefit? We introduce several such methods, based on one oncology trial. For blinding purposes, data had to be manipulated with random error but the general considerations remain untouched.

THE TRIAL

The trial has a simple 2-arm parallel group design with a 1:1 randomization to either the experimental or the control group. Primary efficacy endpoint is disease-free survival (DFS) defined as the time from randomization to disease recurrence, second primary malignancy or death of any cause, whichever occurs first. Overall survival (OS) is an important secondary efficacy endpoint. The crossover option was introduced after an early positive interim analysis demonstrating superiority of the experimental treatment vs. control for DFS. A high percentage of patients in the control arm (more than 30%) decided to cross over. It is important to note that patients in the control arm with prior disease recurrence were not allowed to crossover. This is because the interim results did not allow any conclusion on whether this group of patients would also benefit from experimental treatment.

PRE-PLANNED ANALYSES AND RESULTS

The following analyses were to be conducted at a follow-up analysis after crossover according to the statistical analysis plan, for both DFS and OS:

1. ITT analysis: this analysis is always requested by both authorities and payers and is the most important method for drug approval. Results are conservative in the sense that the "true" treatment effect is likely to be underestimated because it completely ignores that crossover patients will generally benefit from treatment.
2. Sensitivity analysis censoring patients in the control arm at the time of crossover (first administration of experimental treatment): This method is known to be prone to bias as the patient's decision to crossover is usually not independent of prognostics. However, the direction of the bias is usually unclear – it may be in either direction (pro and contra experimental treatment).

The focus in the remainder will be on OS only as the corresponding considerations for DFS are generally very similar.

	Overall survival	
Method	Hazard ratio	95% Confidence interval
ITT	0.80	(0.68, 0.94)
Censored	0.57	(0.48, 0.67)

Table 1: Results of pre-planned ITT and sensitivity analyses (experimental vs. control)

The hazard ratio of 0.80 from the ITT analysis is likely to underestimate the true treatment benefit for overall survival, as outlined above. In contrast, the hazard ratio of 0.57

from the sensitivity analysis is very likely to overestimate the survival benefit in this trial. This is because patients in the control arm with prior recurrence were not allowed to crossover. Thus, censoring control patients at the time of crossover, only removes patients without prior recurrence from the set-at-risk whereas all patients with prior recurrence are kept. Consequently, the set-at-risk in the control arm at later time points is biased to patients with a poorer prognosis for overall survival compared to the overall population.

Can we apply another method to get a better understanding of the true survival benefit, which seems to lie somewhere between 0.57 and 0.80 in terms of the hazard ratio?

FURTHER STATISTICAL METHODS FOR CROSSOVER ADJUSTMENT

On top of the 2 methods outlined above (ITT and censoring), the following other methods are currently known:

1. Naïve methods:
    a) Excluding crossover patients completely from the analysis (creates a non-randomized subset with reduced sample size)
    b) Analyses with a time-varying covariate for treatment or switch

    Both naïve methods assume that there are no confounders i.e. variables influencing both survival and the decision to crossover and are thus not recommended due to a high probability of bias.
2. Complex methods:
    a) Inverse probability of censoring weighting
    b) Rank-preserving structural failure time method
    c) Other less common methods (for an excellent non-technical overview, see [1])

INVERSE PROBABILITY OF CENSORING WEIGHTING (IPCW)

This method uses time-dependent Cox proportional hazard models with time-varying weights, giving larger weight to non-crossover patients who have similar characteristics as patients who crossed over. The crucial assumption is that there are no unmeasured confounders i.e. all factors, influencing both the decision to crossover and survival, are included in the weight calculation. IPCW was not considered suitable for our

trial because there were no prognostic variables assessed during the trial – and it is very unlikely that baseline characteristics alone are sufficient to satisfy the "no unmeasured confounders" assumption.

RANK PRESERVING STRUCTURAL FAILURE TIME (RPSFT) MODELS

This method uses concepts of latent survival time and counterfactual survival time: A patient who crosses over to experimental treatment has a counterfactual survival time – the time-to-death if no experimental treatment had been received. This counterfactual survival time is estimated based on accelerated failure time models [2]. The main assumptions of this method are as follows:

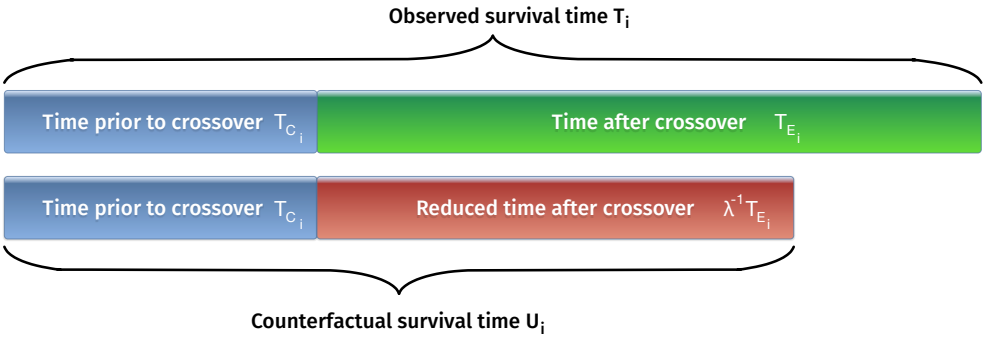
1. Latent survival is the same in both arms (this should be guaranteed via randomization)
2. If a patient x survives longer than patient y with both of them receiving the same treatment, x would also survive longer than y if they received another (but still same) treatment. This is the rank-preserving assumption.
3. Treatment from randomization (at start of study) and later treatment at the time of crossover have the same effect. This is the constant treatment effect assumption.
4. There is a suitable mathematical model relating treatment exposure to survival.

This method was considered suitable for our example trial - due to an early introduction of the crossover option, it seemed sensible to believe that the constant treatment effect assumption is not heavily violated.

A SIMPLE RPSFT MODEL

GENERAL CONCEPT

Principally, the RPSFT model allows for very general and also complicated models relating treatment exposure to survival. However, in practice, usually simple models are applied with the following model being widely used:



This means that we assume that, for a beneficial treatment, there is a suitable factor $\lambda > 1$ (called the "acceleration factor") which prolongs the latent time-to-death by a constant amount from first administration, in all patients who receive the experimental treatment. It is crucial to understand that this assumption does not only hold for the crossover patients but also for all patients in the experimental treatment arm ($T_{C_i} = 0$ and thus $U_i = \lambda^{-1}T_{E_i}$) and the non-crossover patients in the control arm ($T_{E_i} = 0$ and thus $U_i = T_{E_i}$). Furthermore, the actual treatment duration is not taken into account but we assume a constant prolongation, starting with the first administration of the experimental treatment.

ESTIMATE THE ACCELERATION FACTOR

According to the model assumptions, the latent survival time (Ui) is a baseline characteristic and has the same distribution in each treatment arm. The general idea is thus to find the acceleration factor which, when removing the treatment effect according to the model and thus calculating the latent survival time for each patient, leads to no difference anymore between the arms. Practically, an iterative process of searching a grid of possible values is used to estimate the acceleration factor - applying the test statistic for the (preferably primary) null hypothesis [3]. If, for example, the log-rank statistic is used for the primary ITT test, the acceleration factor should lead to a log-rank test statistic of 0 for the comparison of latent survival time between the arms.

Of note, there are also other approaches to estimate the acceleration factor (parametric approach suggested by Branson/Whitehead [4] and hybrid approach suggested by Robins [2]).

HANDLING CENSORED DATA

Generally, we do not observe the event of interest in all patients but we also face censored data. Unfortunately, it is not possible to reduce survival time in the same way

for patients without an event: we would introduce bias because censoring would be dependent on time spent on treatment and thus treatment arm (informative censoring).

For censoring at an administrative cut-off date, the solution is to re-censor all patients (with and without event) at the longest time they could have been observed for latent survival **under each possible treatment history**. This surely means that we loose information: the more beneficial the treatment, the earlier is the re-censoring and the more information is lost.

For censoring prior to the administrative cut-off (as for patients lost-to-follow-up), the literature is not really clear: one idea is to assume random censoring and to treat these censored times in the same way as events.

COMPUTATIONAL APPROACH

Given the acceleration factor (λ) and the time from randomization to the administrative cut-off for each patient (CUT_i), the following calculations need to be performed:

1. Compute counterfactual survival times U_i for patients who crossed over in the control arm. For all patients in the control arm, re-censor at $\lambda^{-1}CUT_i$ (in case of no earlier event or censoring).
2. Compare these control arm data to the experimental treatment arm using standard techniques such as Cox-proportional hazard models, Kaplan-Meier curves and log-rank test.
3. Take the point estimates (e.g. hazard ratios, survival rates) in the usual way.
4. Estimates of uncertainty (e.g. confidence intervals, standard errors) must not be calculated in the usual way – they need to be inflated in order to cover the uncertainty in the estimate of the acceleration factor. Generally, bootstrapping techniques can be used. For the confidence interval of the hazard ratio, another method has been developed by White, which inflates the standard error based on the ITT-test used to estimate the acceleration factor [6][7].

REVISITING THE TRIAL DATA

We come back to our original motivation to better understand the true survival benefit in our trial, which seems to lie somewhere between the hazard ratio of 0.57 (ITT method) and 0.80 (method censoring cross-over patients at the date of first administration of experimental treatment).

Applying the simple RPSFT model, as outlined above, to the trial data, leads to an acceleration factor of 1.43 and the following results for the hazard ratio:

	Overall survival		
Method	Hazard ratio	95% Confidence interval (Bootstrap method for RPSFT)	95% Confidence interval (ITT method for RPSFT)
ITT	0.80	(0.68, 0.94)	
Censored	0.57	(0.48, 0.67)	
RPSFT	0.75	(0.65, 0.89)	(0.61, 0.92)

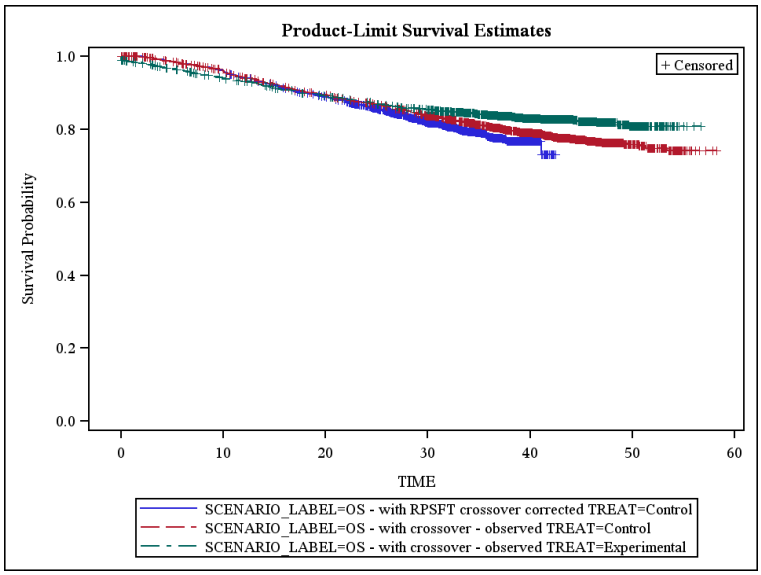
Table 2: Hazard ratios for overall survival (experimental vs. control)

Thus, the values calculated by the RPSFT-model (both for the point estimate and the confidence interval) indeed lie between those from the other 2 analyses and, in fact, are closer to the values from the ITT analysis. The RPSFT-adjusted confidence intervals by bootstrapping and by the White-method do not notably differ.

The RPSFT-adjusted Kaplan-Meier curve for the control arm nicely shows the effect of the artificial re-censoring:

KM survival estimates for RPSFT corrected survival times **1**

The LIFETEST Procedure



CONCLUSION

In trials with a considerable amount of crossover to experimental treatment, the standard ITT method may strongly underestimate the treatment effect. The use of statistical methods to correct for crossover is an important tool to get a better understanding of the true treatment benefit.

Naïve and simple methods to correct for crossover are very prone to bias and generally not recommended. Correspondingly, the use of more sophisticated methods is strongly encouraged. However, caution needs to be taken here as well in order to see which method is best suited for the specifics of each trial - there is currently no single method known which could universally be applied.

REFERENCES

[1] Watkins et al., Adjusting overall survival for treatment switches: commonly used methods and practical application, Pharmaceutical Statistics 2013: 348-357
[2] Robins J, Tsiatis A: Correcting for noncompliance in randomized trials using rank preserving structural failure time models. Communications in Statistics – Theory and Methods 1991; 20: 2609-2631
[3] Robins J, Blevins D, Ritter G, Wulfsohn M.: G-estimation of the effect of prophylaxis therapy for Pneumoycystis carinii pneumonia on the survival of AIDS patients, Epidemiology 1992; 3:318-336
[4] Branson M, Whitehead J: Estimating a treatment effect in survival studies in which patients switch treatment. Statistics in Medicine 2002, 21:2449-2463
[5] White I.: Randomization-based methods for correcting for treatment changes, examples from the Concorde trial. Statistics in Medicine 1999; 18(19):2617-2635
[6] White I, Walker S, Babiker A: strbee: Randomization-based efficacy estimator. The Stata Journal 2002; 2(2): 140-150

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Sabine Lauer
Dr. Lauer Research
Neuhöfer Straße 43a
D-63263 Neu-Isenburg

Work Phone: +49 176 49003254
Email: sabine.lauer@drlauer-research.com
Web: www.drlauer-research.com