Analysis of RCTs in the presence of non-compliance

Constantin T Yiannoutsos, Ph.D.

March 17, 2018

Outline

Causality

Randomization allows us to make causal statements regarding the effect of treatment on the response of interest.

In this lecture we will give a more formal definition of cause and effect. To do this we use what are called *counterfactual* random variables.

As usual, we consider an overall population of individuals that we are interested in and assume that the participants in a clinical trial represent a random sample from this population. Within this clinical trial we will compare an experimental treatment (e.g., treatment 1) to a standard treatment or placebo (treatment 0).

Definition of counterfactual random variables

We define the counterfactual random variable Y_1^* to denote the response. This may be a binary or continuous outcome that a randomly selected individual would have if, possibly contrary to fact, they received treatment 1 (the experimental treatment).

Similarly, we define the counterfactual random variable Y_0^* to denote the response that a randomly selected individual would have if, possibly contrary to fact, that individual received treatment 0 (standard treatment or placebo).

We imagine that both random variables Y_0^* and Y_1^* exist, even though in actuality it would be impossible to observe both responses on the same individual (from which the term *counterfactual* or "contrary to fact" emanates).

Formal definition of the causal effect of treatment

Definition 1.1 (Causal treatment effect).

At the individual level, we say that treatment *causes* the effect

$$Y_1^* - Y_0^*$$

The mean causal effect of treatment

Clearly, if we knew the response of an individual to both treatments, then we would choose whichever treatment gave the better response.

Of course, this is not possible at the individual level but perhaps we can look at this question at the population level. That is, we will estimate the causal treatment response by the population mean causal effect

$$\Delta = E(Y_1^* - Y_0^*) = E(Y_1^*) - E(Y_0^*)$$

If Δ is positive, then, on average, the response on treatment 1 will be better than on treatment 0.

Note that, at the individual level, this does not necessarily imply that any specific individual will be guaranteed to benefit from the treatment found to be superior based on Δ but, on average, the population as a whole will benefit.

Observable data

The data that we actually observe from a clinical trial are summarized by (Y_i, A_i, X_i), with $i = 1, \cdots, n$, where, for a randomly selected individual i,

- $A_i = (0, 1)$ denotes the treatment assignment (to the new treatment or the standard treatment or placebo respectively)
- Y_i denotes the response
- X_i denotes any additional characteristics, collected on the individual prior to treatment assignment (baseline characteristics)

We will refer to these as the observable random variables.

Note: We distinguish between the *observed response* Y_i for the i-th individual and the *counterfactual responses* Y_{1i}^* and Y_{0i}^*.

The consistency assumption

We make the reasonable assumption that $Y_i = Y_{1i}^*$ if $A_i = 1$ and that $Y_i = Y_{0i}^*$ if $A_i = 0$. In other words, we assume that the observed treatment response is equal to the counterfactual treatment response if the individual were assigned the same treatment as the one we observe to be assigned to the individual. This is the *consistency* assumption.

Assumption 1.1 (Consistency).

The consistency assumption in causal models is defined as

$$Y_i = Y_{1i}^* I(A_i = 1) + Y_{0i}^* I(A_i = 0)$$

where $I(\cdot)$ denotes the indicator function of an event and it equals 1 when the event is true and 0 otherwise.

Association versus causation

Traditional statistical methods allow us to make *associational* relationships. For example, we can use regression models that allow us to estimate relationships such as $E(Y_i|A_i, X_i)$. These models explore the association of the outcome Y_i, for the i-th subject, to the assigned treatment A_i and other measured characteristics (prognostic factors or covariates) X_i.

These associational relationships are not the causal relationships that are the parameters of interest.

However, associational statements are more easily assessed. So the question of estimating causal effects is modified as:

"Under what conditions or assumptions can we estimate causal parameters such as Δ, from observable data?"

Randomization

This is where randomization plays a key role. Since treatment is randomly assigned to the patient in a randomized study, treatment assignment is independent of any pre-treatment characteristics of the individual.

Consequently, we make the following assumption:

Assumption 1.2 (Independence).

In randomized clinical trials,

$$A_i \text{ is independent of } (Y_{1i}^*, Y_{0i}^*, X_i)$$

That is, randomization severs any association between how an individual would have responded if given treatment 1 and how he/she would have responded if given treatment 0 and the treatment he/she was randomized to.

Remark

It is important to note that the assumption of independence between the treatment assignment A_i and the counterfactual response of individual i, (i.e., Y_{1i}^* or Y_{0i}^*), is not the same as saying that A_i is independent of Y_i (the observed response).

Since $Y_i = Y_{1i}^* I(A_i = 1) + Y_{0i}^* I(A_i = 0)$, Y_i is a function both of counterfactual responses and the treatment assignment and, as such, will not be independent of A_i.

In fact, if treatment is effective, as one hopes, then we would expect (and want) Y_i to depend on A_i.

Implication of randomization

We will now use assumptions (1.1) and (1.2) to show that

$$P(Y_{1i}^* \leq u)^1 = P(Y_i \leq u | A_i = 1)$$

This follows because

$$P(Y_i \leq u | A_i = 1) = P(Y_{1i}^* \leq u | A_i = 1)$$

by the consistency assumption (1.1). In addition,

$$P(Y_{1i}^* \leq u | A_i = 1) = P(Y_{1i}^* \leq u)$$

which is a consequence of the independence assumption (1.2).

Similarly, we can show that $P(Y_{0i}^* \leq u) = P(Y_i \leq u | A_i = 0)$.

[1] $P(Y_i \leq u)$ is a fancy way of symbolizing the distribution of probability of the response Y_i being below a number u. For example, this could be the probability of response to cancer treatment being at most 20%.

Implications of randomization

From the previous relations, the average causal treatment effect

$$\begin{aligned} \Delta &= E(Y_1^*) - E(Y_0^*) \\ &= E(Y|A=1) - E(Y|A=0) \end{aligned}$$

Now we have an expression for the causal parameter Δ in terms of quantities that can be estimated.

Estimation of the causal effect

To estimate Δ it suffices to estimate $E(Y|A=1)$ and $E(Y|A=0)$.

These can be estimated by $\bar{Y}_1 = \sum_{i=1}^{n} Y_i I(A_i = 1)/n_1$ and $\bar{Y}_0 = \sum_{i=1}^{n} Y_i I(A_i = 0)/n_0$ respectively, where n_1 and n_0 are the treatment-specific sample sizes.

Thus, an unbiased estimator for the causal treatment effect Δ can be derived from a randomized study using

$$\hat{\Delta} = \bar{Y}_1 - \bar{Y}_0$$

Non-compliance

The arguments outlined above assume that patients take the treatment to which they are randomized. In most clinical trials however, this is rarely the case. This is called *non-compliance*.

There is almost always some form of noncompliance from the intended treatment regimen. Some reasons for non-compliance are:

- A refusal by the patient to start or continue the assigned treatment, due to side effects or a belief that the treatment is ineffective
- A failure to comply with detailed instructions, such as drug dose, or to attend examinations when requested to do so
- A change of treatment imposed by the physician for clinical reasons, such as adverse effects or deterioration of the patients health
- An administrative error. In its most extreme form, this may be the implementation of the wrong treatment.

Analytical strategies in the face of non-compliance

Some strategies that have been proposed include the following:

- **Intent-to-Treat Analysis** (ITT; As randomized)
  Everyone is included in the analysis and the comparison of treatments is based on the difference of the average response between the randomized groups ignoring the fact that some patients were non-compliant.
- **As-treated analysis**
  This type of analysis follows the general idea that only patients who fully complied with their assigned treatment regimen are to be compared and all non-compliant patients are excluded from the analysis.

The ITT principle and the dogma of clinical trials

The intent-to-treat (ITT) analysis principal complies (no pun intended) with the central dogma in clinical trial research:

Exclusions based on post-randomization considerations, such as noncompliance, are not allowed for the primary analysis.

This is because exclusion of patients from the analysis may result in bias in the treatment comparisons.

Example: The Clofibrate study

To illustrate some of the difficulties that can result from non-compliance, we consider the results from a study conducted by the Coronary Drug Project (New England Journal of Medicine, 1980).

This was a double-blind placebo-controlled trial comparing Clofibrate to Placebo. The following table shows the results from the ITT analysis:

Table 1: Intent-to-Treat Analysis

	Treatment	
	Clofibrate $N = 1065$	Placebo $N = 2695$
5-year mortality	0.18	0.19

Effect of non-compliance

Table 2: 5-year mortality by treatment adherence

	Clofibrate		Placebo	
Adherence (% of capsules taken)	5-year mortality	Number of patients	5-year mortality	Number of patients
Poor ($< 80\%$)	0.25	357	0.28	882
Good ($> 80\%$)	0.15	708	0.15	1813

It is clear from these data that compliant patients are prognostically different from non-compliant patients. Therefore, the as-treated approach may lead to severe biases because it cannot separate the prognostic effect of noncompliance from the prognostic effect of treatment.

By contrast, the intent-to-treat analysis does not suffer from this type of bias. At the same time, when some patients do not comply with the treatment, an ITT analysis would diminish the effect of a treatment.

A simple trial design

Consider a randomized study where patients are randomized with equal probability to active drug (treatment 1) or placebo (control) (treatment 0).

Response is dichotomous. The main goal of the clinical trial is to estimate the difference in the probability of response between active drug and placebo

For simplicity, we assume that every patient either takes their assigned treatment or not (partial compliance is not considered) and their compliance can be assessed by a simple assay.

We also consider that the patients assigned to placebo do not have access to the study drug and that compliance cannot be determined for these patients.

Counterfactual and observable random variables

The problem above can be conceptualized as follows:

Let the counterfactual random variables Y_1^* and Y_0^* denote the response (1=response, 0=non-response) of a randomly selected individual if they received treatment 1 or 0 respectively.

Also let C denote the counterfactual random variable corresponding to whether or not a randomly selected individual complies or not $C = (1, 0)$. This is a *counterfactual* random variable because we do not know the compliance status for patients randomized to placebo.

Counterfactual random variables (continued)

Denote by $\theta = P(C = 1)$ the population probability of complying with the assigned treatment, while $\pi_1^{\mathrm{COM}} = P(Y_1^* = 1|C = 1)$ and $\pi_1^{\mathrm{NC}} = P(Y_1^* = 1|C = 0)$ are the probability of response among those who comply or do not comply if given active drug respectively.

Also, denote by $\pi_0^{\mathrm{COM}} = P(Y_0^* = 1|C = 0)$ and $\pi_0^{\mathrm{NC}} = P(Y_0^* = 1|C = 0)$ the probability of response among those who comply or do not comply if given active placebo.

As it is not reasonable to assume that Y_1^* and Y_0^* are independent of C, so we would not expect $\pi_1^{\mathrm{COM}} = \pi_1^{\mathrm{NC}}$ or $\pi_0^{\mathrm{COM}} = \pi_0^{\mathrm{NC}}$.

Estimating the causal treatment effect

Using some simple probability calculations we get that

$$\begin{aligned} E(Y_1^*) &= P(Y_1^* = 1) \\ &= P(Y_1^* = 1|C = 1)P(C = 1) + P(Y_1^* = 1|C = 0)P(C = 0) \\ &= \pi_1^{\mathrm{COM}}\theta + \pi_1^{\mathrm{NC}}(1-\theta) = \pi_1 \end{aligned}$$

and, similarly,

$$E(Y_0^*) = P(Y_0^* = 1) = \pi_0^{\mathrm{COM}}\theta + \pi_0^{\mathrm{NC}}(1-\theta) = \pi_0$$

Therefore, the average causal treatment effect equals

$$\Delta = E(Y_1^*) - E(Y_0^*) = \pi_1 - \pi_0 = \Delta^{\mathrm{COM}}\theta + \Delta^{\mathrm{NC}}(1-\theta)$$

where $\Delta^{\mathrm{COM}} = \pi_1^{\mathrm{COM}} - \pi_0^{\mathrm{COM}}$ and $\Delta^{\mathrm{NC}} = \pi_1^{\mathrm{NC}} - \pi_0^{\mathrm{NC}}$.

Estimation of the counterfactual probabilities π_1^{COM}, π_0^{COM}

How can we estimate the counterfactual probabilities π_1^{COM}, π_0^{COM} using the data we observe from a randomized clinical trial when there is noncompliance?

- θ (the overall compliance rate)

$$P(C = 1|A = 1) = P(C = 1) = \theta$$

- π_0 (the overall response rate)

$$P(Y = 1|A = 0) = P(Y_0^* = 1|A = 0) = P(Y_0^* = 1) = \pi_0$$

- π_0^{NC} (the response rate in the control arm under non-compliance)

$$\begin{aligned} P(Y = 1|A = 1, C = 0) &= P(Y_0^* = 1|A = 1, C = 0) \\ &= P(Y_0^* = 1|C = 0) = \pi_0^{\mathrm{NC}} \end{aligned}$$

- π_1^{COM} (the response rate in the active arm under compliance)

$$\begin{aligned} P(Y = 1|A = 1, C = 1) &= P(Y_1^* = 1|A = 1, C = 1) \\ &= P(Y_1^* = 1|C = 1) = \pi_1^{\mathrm{COM}} \end{aligned}$$

- π_0^{COM} (the response rate in the control arm under compliance)

$$\pi_0^{\mathrm{COM}} = \frac{\pi_0 \pi_0^{\mathrm{NC}}(1-\theta)}{\theta}$$

Estimation of the treatment effect under Intent-to-treat

We are estimating

$$\Delta_{\mathrm{ITT}} = P(Y = 1|A = 1) - p(Y = 1|A = 0)$$

Again, by the assumptions made and some probability calculations we get

$$P(Y = 1|A = 1) = \pi_1^{\mathrm{COM}}\theta + \pi_0^{\mathrm{NC}}(1 - \theta)$$

Similarly,

$$P(Y = 1|A = 0) = \pi^{\mathrm{COM}}\theta + \pi_0^{\mathrm{NC}}(1 - \theta)$$

Thus, Δ_{ITT} is an estimate of

$$\begin{aligned} & \pi_1^{\mathrm{COM}}\theta + \pi_0^{\mathrm{NC}})(1 - \theta) - \left\{\pi_0^{\mathrm{COM}}\theta + \pi_0^{\mathrm{NC}})(1 - \theta)\right\} \\ = \quad & \theta\Delta^{\mathrm{COM}} \end{aligned}$$

This, unless the compliance rate is 100%, will be less (and perhaps much less) than the causal treatment effect under full compliance.

Remarks

Recall that

$$\Delta^{\mathrm{COM}} = P(Y_1^* = 1|C = 1) - P(Y_0^* = 1|C = 1) = E(Y_1^* - Y_0^*)$$

is the difference in the mean counterfactual responses between the two treatment arms among patients that would comply with treatment.

As such, Δ^{COM}, some argue, is the causal parameter of greatest interest since it quantifies the benefit among patients who will comply with the new treatment.

However, we are in fact able to estimate Δ^{COM} since we can estimate the parameter θ, the overall compliance rate if offered the new treatment by

$$\Delta^{\mathrm{COM}} = \frac{P(Y = 1|A = 1)P(Y = 1|A = 0)}{P(C = 1|A = 1)}$$

Since all the quantities are easily estimated from the data of a clinical trial, this means we can estimate the causal parameter Δ^{COM}.

More remarks on the ITT principle

If the null hypothesis of no treatment effect is true; namely

$$H_0 : \ \Delta^{\mathrm{COM}} = \Delta^{\mathrm{NC}} = \Delta = 0$$

the intent-to-treat analysis (which estimates $\Delta^{\mathrm{COM}}\theta$) gives an unbiased estimator of treatment difference (under H_0) and can be used to compute a valid test of the null hypothesis

However, the above results make it clear that the ITT analysis will yield an estimator which diminishes a causal treatment effect under compliance.

As-treated analysis

In one version of an as-treated analysis we compare the response rate of patients randomized to active drug who comply to all patients randomized to receive the control. That is, we compute

$$\Delta_{\mathrm{AT}} = \bar{Y}_{A=1,C=1} - \bar{Y}_{A=0}$$

After some algebra we get that

$$\Delta_{\mathrm{AT}} = \Delta + (1-\theta)(\pi_1^{\mathrm{COM}} - \pi_1^{\mathrm{NC}})$$

where Δ is the average causal treatment effect.

This makes clear that when there is noncompliance, (i.e., when $\theta < 1$), the as-treated analysis will yield an unbiased estimate of the average causal treatment effect only if $\pi_1^{\mathrm{COM}} = \pi_1^{\mathrm{NC}}$.

Since this assumption is not generally true, the as-treated analysis can result in biased estimation even under the null hypothesis.

Cross-over and non-compliance in time-to-event studies

Cross-over from or to active treatment at toxicity or disease progression may lead to statistical challenges in the analysis of overall survival because crossover leads to information loss and dilution of comparative clinical efficacy.

Cross-over (as well as other forms of non-compliance) has potentially significant implications for estimates of survival (e.g., hazard ratios).

The following follows the article by Jönsson and colleagues (Value in Health, 2014), where various methods to account for non-compliance and cross-over are reviewed.

The overall conclusion is that the results of the statistical analysis are potentially very different depending on the method used.

Case study: Sunitimib

We will describe two studies of the drug sunitinib (Sutent; Pfizer, Inc., New York, NY), an orally administered, multitargeted tyrosine kinase inhibitor.

We described here two trials of sunitinib:

1. The study by Motzer et al., (NEJM,) of sunitinib for the treatment of metastatic renal cell carcinoma (mRCC)
2. The study by Demetri et al., (Lancet,) of sunitinib for imatinib-resistant gastrointestinal stromal tumors (GIST)

The issue that complicates routine ITT analyses in both studies is the structured cross-over from the non-sunitinib arm (interferon-alpha and placebo respectively) to sunitinib anticipated by the trial design.

Analytical strategies in the sunitinib trias

There are four analytical strategies that were considered in this study:

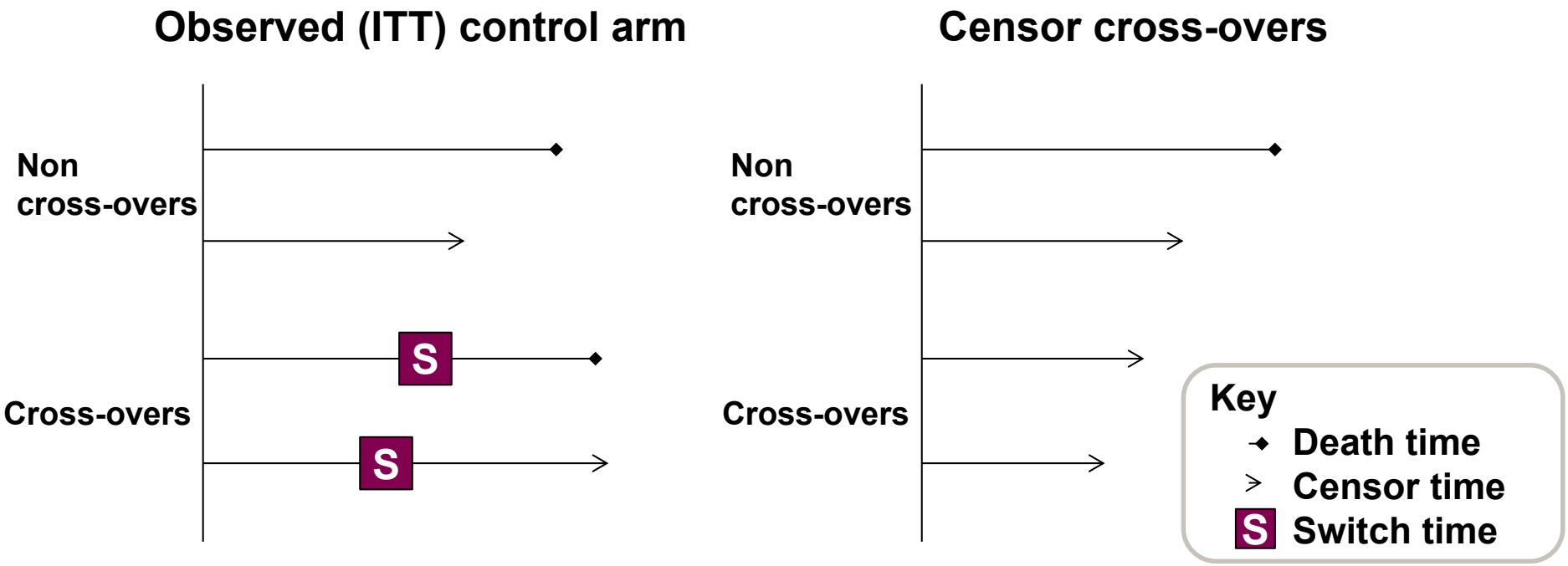
- The intent-to-treat (ITT) or as-randomized approach
- Censoring subjects at the time of cross-over (on-treatment) analysis
- Inverse probability of censoring weights (IPCW) modeling
- Rank-preserving structural failure time (RPSFT) model.

The ITT and on-treatment analyses

The ITT analysis considers everyone's outcome within their randomly assigned treatment. Based on what we saw earlier, the ITT analysis will likely underestimate the treatment effect in the presence of crossover.

The on-treatment approach censors the time of a patient at their cross-over time. The two methods are shown graphically in Figure 1.

Figure 1: ITT versus on-treatment analyses



The IPCW model[2]

In the Inverse probability of censoring weights (IPCW), the model addresses the counterfactual question of what would be the treatment effect in the absence of cross-over. Similarly, it can answer the question with "cross-over" replaced by "non-compliance".

In this method, patients who cross over (or exhibit non-compliance) are censored, while patients remaining in their randomized arm (or continuing to exhibit compliance with their treatment) are weighted to compensate for missing data.

The weights are determined by the predicted probability of not being censored at a given time. Then, a survival analysis is carried out with weights made up of the inverse of the probability of remaining uncensored based on each patient's profile (measured covariates both at baseline and obtained over time during the study).

[2] Robins & Finkelstein, 2000.

Constructing the weights

The IPCW weights are constructed in two steps:

1. Step 1 Calculate the probability of cross-over at each time point based on baseline characteristics only; these are the: *numerator* weights
2. Step 2 Calculate the probability of cross-over at each time point based on both baseline and time-updated characteristics: *denominator* weights

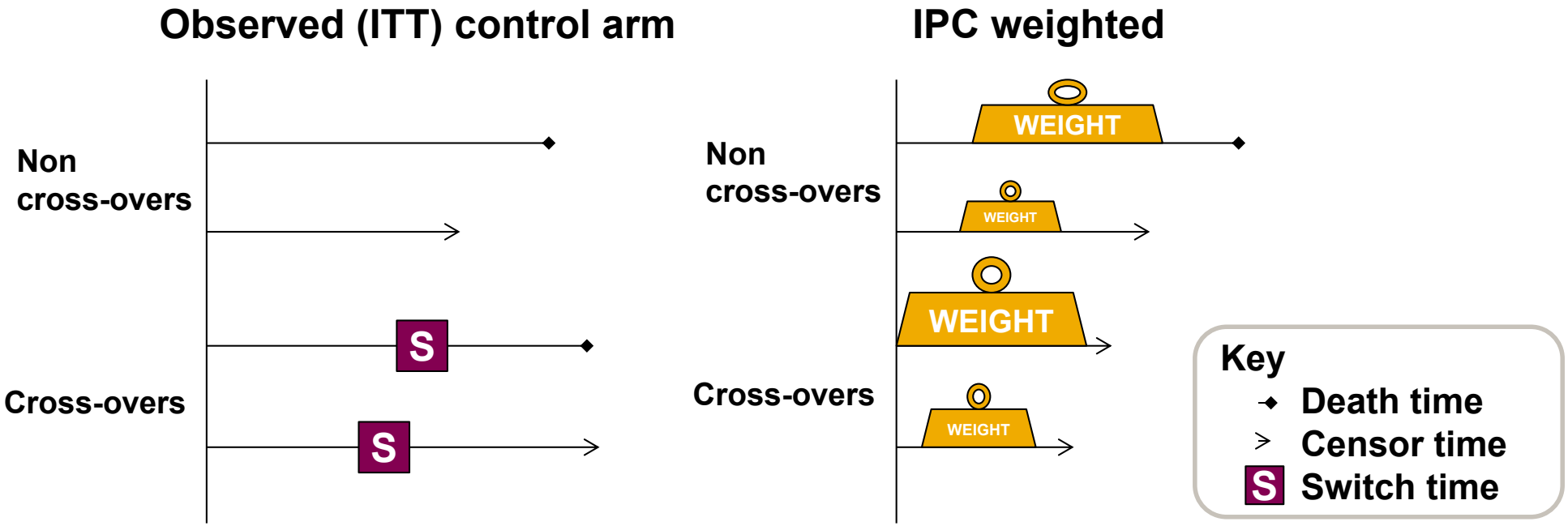
IPCW (stabilized) weights are calculated as the ratio of the numerator over the denominator weights.

A weighted analysis of the usual survival model (e.g., Cox, parametric models, etc.) is run. However, the interpretation concerns the (possibly counterfactual) *policy* of no cross over.

Schematic representation of the IPCW method

Overall, patients who belong to groups with high cross-over rate will get higher weights (they will "represent" more patients who crossed over) while patients in groups with low cross-over rate will get lower weights (they will represent a smaller number of patients who crossed over). This is shown schematically in Figure 2.

Figure 2: Schematic representation of the IPCW procedure



The RPSFT model

The Rank-preserving structural failure time (RPSFT) model allows a direct comparison of the two (or more) randomization groups by adjusting the overall survival of patients who cross over to reflect the survival they would have had if they never received the experimental treatment.

In the RPSFT model, the observed failure time T_i for each subject i is associated with the counterfactual time U_i that would have been observed had the subject not crossed over and received the active drug. T_i and U_i are related through the treatment history $Q_i(u)$, $u \in (0, T_i)$ as follows:

$$U_i = \int_0^{T_i} \exp\{\psi Q_i(u)\}\, du$$

If switching occurs at discrete time points, this reduces to the following sum:

$$U_i = \begin{cases} T_i & \text{non} - \text{cross} - \text{overs} \\ T_i^c + \exp(\psi)\, T_i^e & \text{cross} - \text{overs} \end{cases}$$

T_i^e and T_i^c are the times spent in the experimental and control arms respectively and note that $T_i = T_i^c + T_i^e$.

The RPSFT model as an accelerated failure-time model

In the RPSFT model, the coefficient ψ *accelerates* the consumption of the survival time by a factor e^{ψ}. When $\psi < 0$ the untreated survival time U_i is less than T_i, the observed survival, and the treatment is beneficial; otherwise the treatment is detrimental.

The RPSFT model is related to the so-called "accelerated failure time models" in survival analysis where risk factors act multiplicatively on the time scale (i.e., by accelerating or decelerating the time until the occurrence of the failure).

The RPSFT model: An example

To understand how this works, consider a study in smoking cessation[3] resulting in $\psi = -0.1$ and an observed lifetime for a subject $T_i = 2.2$ years, given smoking history 0.2 years on, one year off, one year on. The survival if always smoked is

$$U_i = (2.2 - 2)e^{-0.1} + (2 - 1)e^{0} + (1)e^{-0.1} = 1 + 1.2e^{-0.1} \approx 2.09$$

This means that, had the subject smoked for the entire 2.2 years, their survival would have been $U_i = 2.09$ years.

However, the one-year smoking cessation added about 0.11 years of life to this subject's survival, since his observed survival time was $T_i = 2.2$ years.

Another way to see this is to consider what would happen if the subject permanently quit smoking. In that case, $U_i = (2.2)e^{-0.1} = 1.99$ and $e^{0.1} - 1 = \frac{2.2-1.99}{1.99} = 10.6\%$, the fractional increase in survival due to smoking cessation.

[3] Mark & Robins, 1993.

Core idea of the RPSFT model

The main idea of the RPSFT model is that each patient has an inherent failure time U_i, which, because of randomization, is independent of the treatment assignment R_i.

In other words, people with longer or shorter survival don't end up preferentially in one or the other treatment arm; randomization guarantees this (at least in expectation).

In addition, the decision to cross over from the control to the experimental arm is assumed to be independent of the true failure time if unexposed (in our case if not crossing over) U_i.

Note that this does not mean that the decision of crossing over is independent of T_i, the observed failure time.

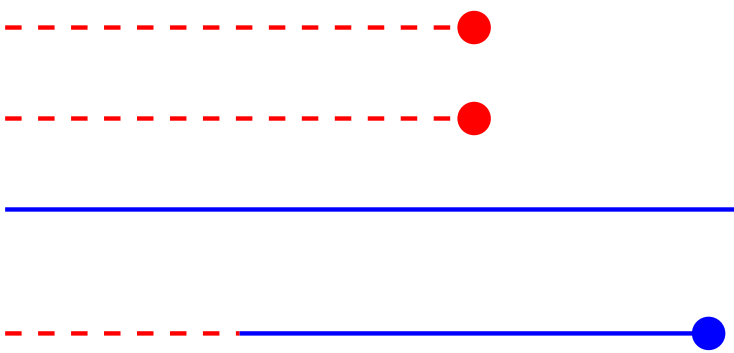
Interpretation of ψ

From the previous discussion, the coefficient ψ is related to the exposed and unexposed time by the following equation

$$e^{\psi} = \frac{T^c}{T^e} \Rightarrow e^{\psi} T^e = T^c$$

That is, $\psi < 0$ results in a longer survival if treated compared to untreated (treatment beneficial).

For example, if $e^{\psi} = 0.5$ (i.e., $\psi \approx -0.69$), this means that one year unexposed/untreated equals two years exposed/treated.



U_i for subject i

Untreated, life = 1 year

Always treated, life = 2 years

Untreated 6 months
then switch, life = 1.5 years

Construction of the RPSFT model

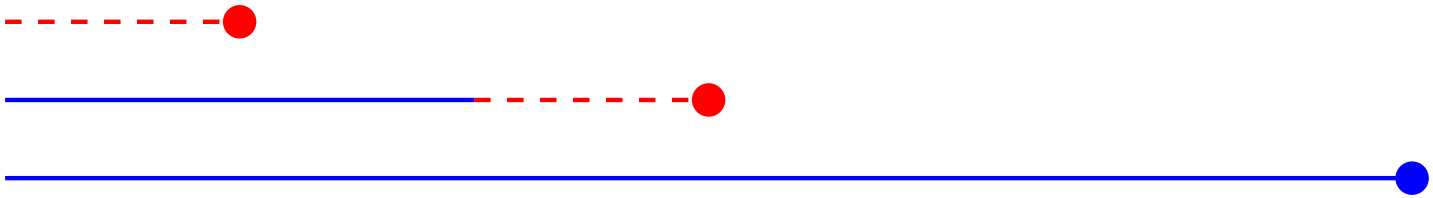
The RPSFT model is constructed in the following steps:

1. Define a model relating the observed event time T_i to the unobserved event time U_i *that would have been observed if crossover had not occurred*.
2. Compute U_i for a range of possible values of ψ (which includes all relevant confounders) and find the one for which a statistical test of the equality of U_i across the two groups has the highest (least significant) p value. This ensures that the U_i are independent of R_i.

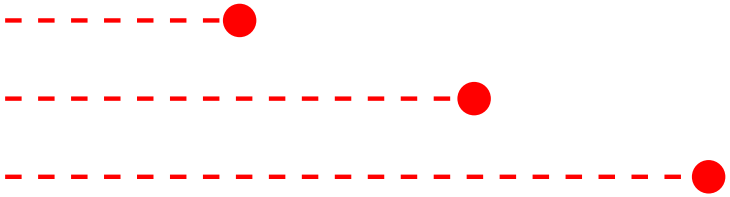
Hypothetical data example[5]: Switches only from active arm.

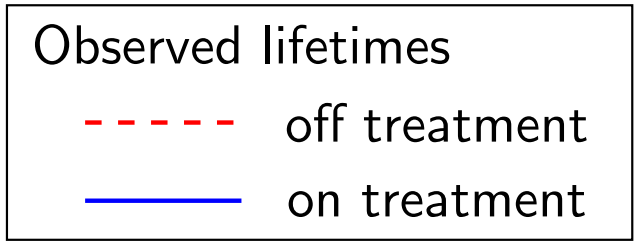
Treated and untreated subjects have equal U_i.[4]

Treated arm:



Control arm:



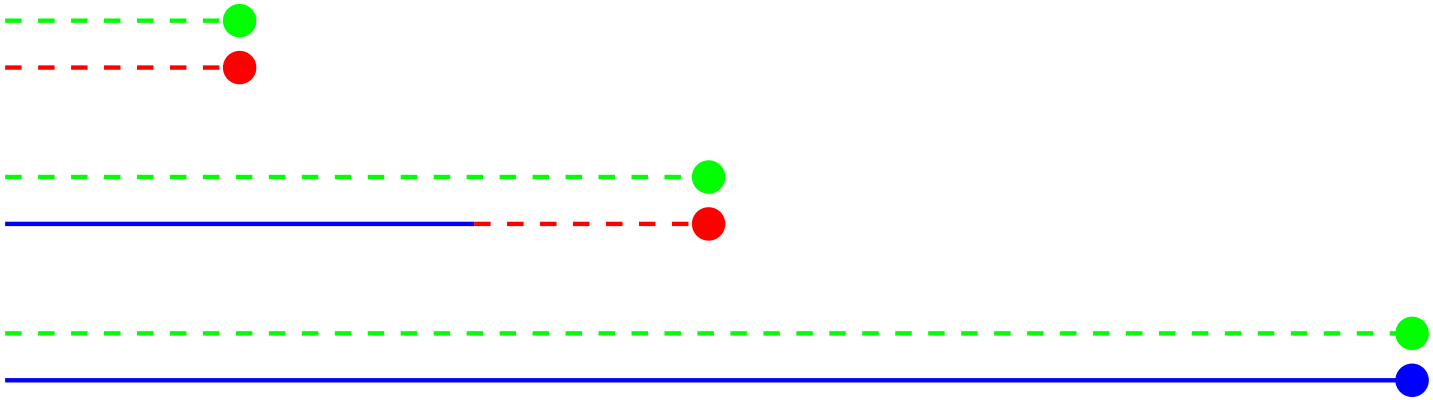


[4] Note that this is due to the randomization.

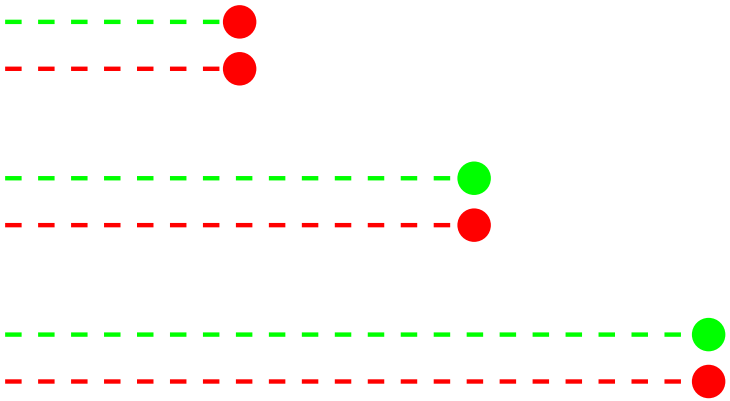
[5] Closely following the presentation by Ian White, HTMR network workshop on Methods for adjusting for treatment switches in late-stage cancer trials. London, 20th February 2012

Hypothetical data example: $e^{\psi} = 1$

Treated arm:



Control arm:



Observed lifetimes

off treatment

on treatment

Fitted untreated lifetimes

Comments

The estimate of ψ resulting in $e^{\psi} = 1$ (i.e., $\hat{\psi} = 0$) above, is not a good estimate for ψ since the estimated untreated survival times are not equal between the two groups!

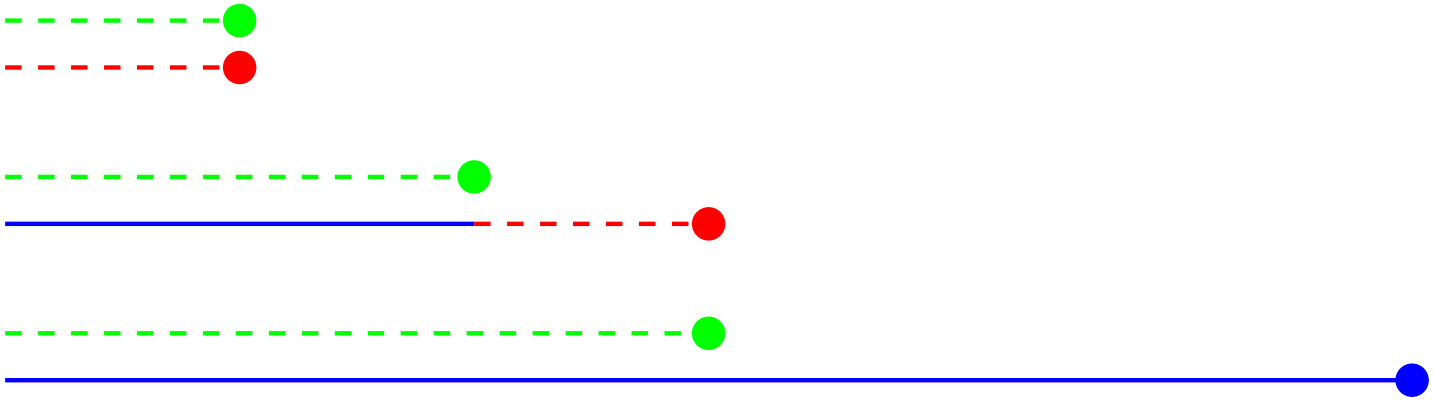
Note that, because of randomization, the estimated untreated survival times must be equal in the two groups. Since the estimate $\hat{\psi} = 0$ resulted in unequal survival times between the two groups, this estimate does not reflect the data.

Note also that $\hat{\psi} = 0$ is equivalent to no treatment effect since then $e^{\psi} = \frac{T^c}{T^e} = 1$. In other words, the survival under exposure would be equal to the survival under non-exposure.

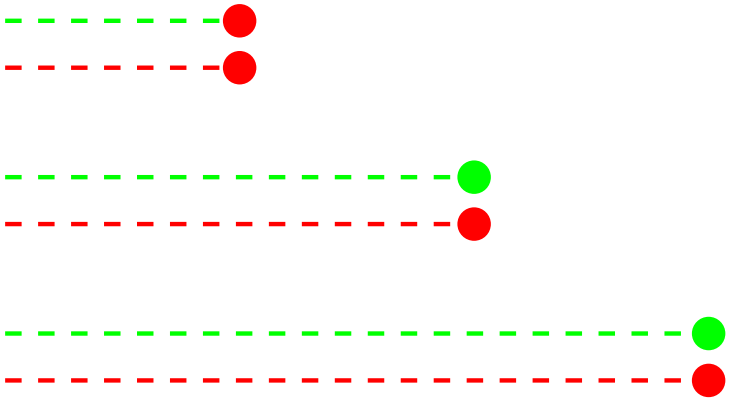
In the above situation, the assumption of no treatment effect is incongruous with what we see in practice, i.e., generally longer observed survival times among the treated patients compared to the untreated patients.

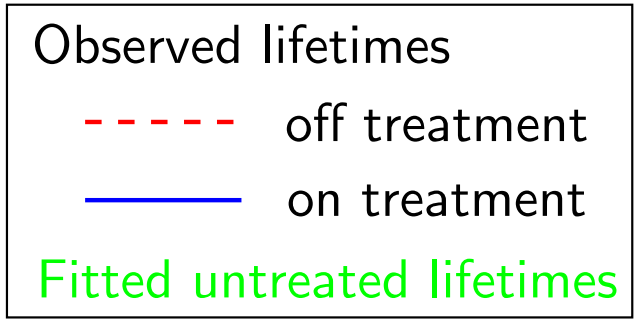
Hypothetical data example: $e^{\psi} = 0.5$

Treated arm:



Control arm:





Thus, $e^{\psi} = 0.5$ (i.e., $\psi \approx -0.69$) is a good estimate for ψ since the estimated untreated survival times are balanced between the two groups.

Recensoring

Censoring introduces unexpected complications into the RSFMT model.

This is because, if there is a beneficial treatment effect (which extends survival), then failure times in the treated group will be more likely to be censored. Thus, censoring is informative and excluding censored observations will negatively bias the estimate of the treatment effect[6].

Mathematically, censoring implies that, instead of the failure time T_i we observe $X_i = \min(T_i, C_i)$ where C_i is the censoring time, which runs from randomization to the common closure of the study.

Unfortunately, replacing U_i with X_i in the calculations will not work unless the null hypothesis is true (i.e., if $\psi = 0$) because if non-compliance is non-random, X_i and R_i are not independent from each other.

[6] The exact opposite happens when a treatment or exposure is detrimental.

Dealing with censoring in the RPSFT model

The core argument for developing a procedure to deal with censoring is to understand that, just as with U_i, C_i (the maximum follow-up time) is (at least in theory) known at randomization and is thus independent of R_i the randomly assigned treatment.

Thus, any function of U_i and C_i will be independent of R_i as well. We define such a function as $X_i(\psi) = \min(U_i(\psi), C_i(\psi))$ as follows:

$$C_i(\psi) = \begin{cases} C_i, & \text{if} \psi \geq 0 \\ C_i \exp(\psi) & \text{if} \psi < 0 \end{cases}$$

where we have made the dependence on ψ explicit. This is called *recensoring* the data.

If the treatment is beneficial, $C_i(\psi)$ is the censoring time *that would have been observed* under no treatment effect. Otherwise, $C_i(\psi)$ is the censoring time under no effect. The new censoring indicator is

$$\Delta_i(\psi) = I\{C_i(\psi) < U_i(\psi)\}$$

We say then that the ith individual is ψ-censored.

Statistical implementation of the RPSFT model

Using the fundamental equation

$$U_i = \int_0^{T_i} \exp\left[\psi Q_i(u)\right] du$$

we can estimate $U_i(\psi)$ for a specific value of ψ.

Then, considering these $U_i(\psi)$ as the true failure times, we carry out a statistical test (e.g., a log-rank test) of the treatment arms.

The value ψ_0 we seek is the solution satisfying the equation[7]

$$\Pr(U_i(\psi_0) \geq x | R_i = 1) = \Pr(U_i(\psi_0) \geq x | R_i = 0)$$

Thus, the desired value of ψ_0 is that which results in the most non-significant log-rank (or other statsitical) test.

[7] Note that this is the definition of independence between U_i and R_i.

Practical implementation of the RPSFT model

In practice, we carry a grid search to find this value of ψ_0.

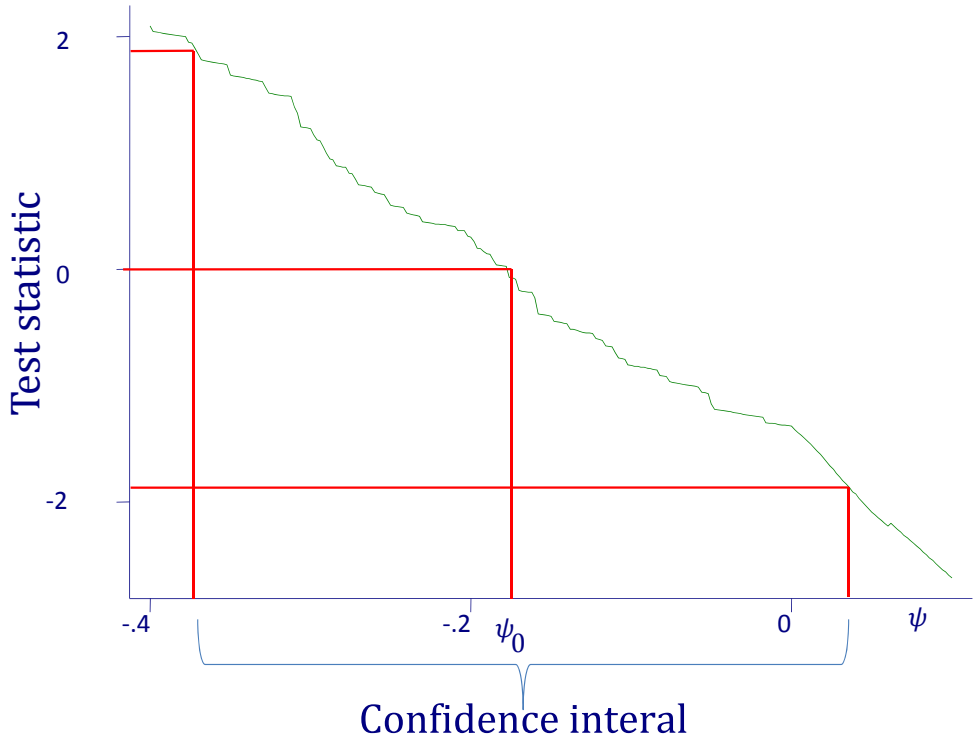


Figure 3: Hypothetical grid search for ψ_0

The confidence interval is the region where the test statistic does not reject the null hypothesis.

Remarks on the RPSFT model

The RPSFT model is "rank preserving" because a constant factor is used for adjusting the time to event for each patient. Thus, if two patients i and j are on the same treatment (either control or experimental), and patient i fails (dies) before patient j, before adjustment, patient i will also *always* fail before patient j after adjustment.

In other words, the ranking in failure times is preserved.

A key assumption of the RPSFT model is that the experimental treatment results in a constant change (reduction) in the time to failure or death, which is assumed equal for all patients before and after progression. This may not be a reasonable assumption in some cases, which may restrict the use of the method.

Advantages and disadvantages of the RPSFT model

A major advantage of the RPSFMT model is that it is "randomization respecting" method. In other words, it compares the two treatment groups as they were randomized.

Another major advantage is that, unlike the IPCW model, the RPSFT model does not assume no unmeasured confounding. In other words, it does not assume that we have accounted for the effect of all factors that are associated with both the outcome and the non-compliance.

A disadvantage of the method is the need to re-censor the data.

Another major disadvantage is the assumption that the treatment effect is constant regardless when, in the disease progression, the treatment is applied. This assumption in particular is unlikely to be 100% correct.

Sunitinib for metastatic renal-cell carcinoma (mRCC)

In this international phase III trial[8], 750 patients with mRCC were randomized to receive either sunitinib (n = 375) or interferon-alfa (IFN-α; n = 375).

Crossover was allowed only after an interim analysis had concluded a significant gain in the primary endpoint PFS.

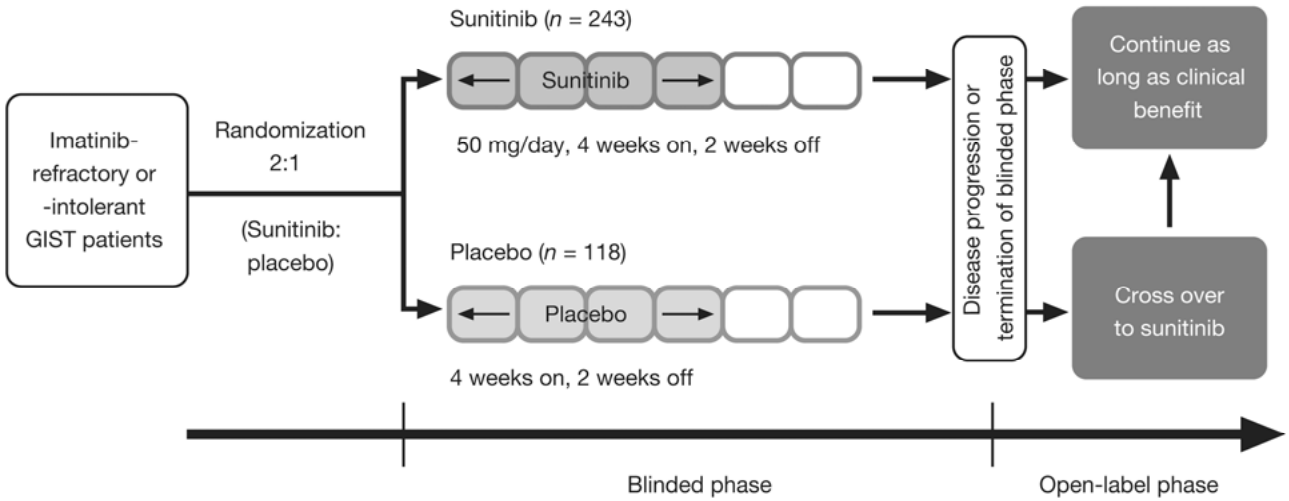
Twenty five patients (7%) in the IFN-α group crossed over to sunitinib after an average of 70.8 weeks. There were 390 total deaths (190 in the sunitinib and 200 in the IFN-α arm).

All events were included in the ITT analysis. Censoring at the time of crossover (on-treatment analysis) led to the exclusion of five deaths in the IFN-α arm which occurred after crossover to sunitinib.

[8] Motzer et al., NEJM, 2007.

mRCC study design

Schematically, the design of the study by Motzer and colleagues is shown in the following Figure:



!h

Figure 4: mRCC study design

Analysis by RPSFT

In the RPSFT model, the estimated value for the acceleration parameter calculated using a grid search method was $\hat{\psi} = -0.244$, corresponding to a decrease in overall survival time of $\exp(\hat{\psi}) = 0.22$ with IFN-α than with sunitinib.

The results of all analyses are presented in the following Figure:

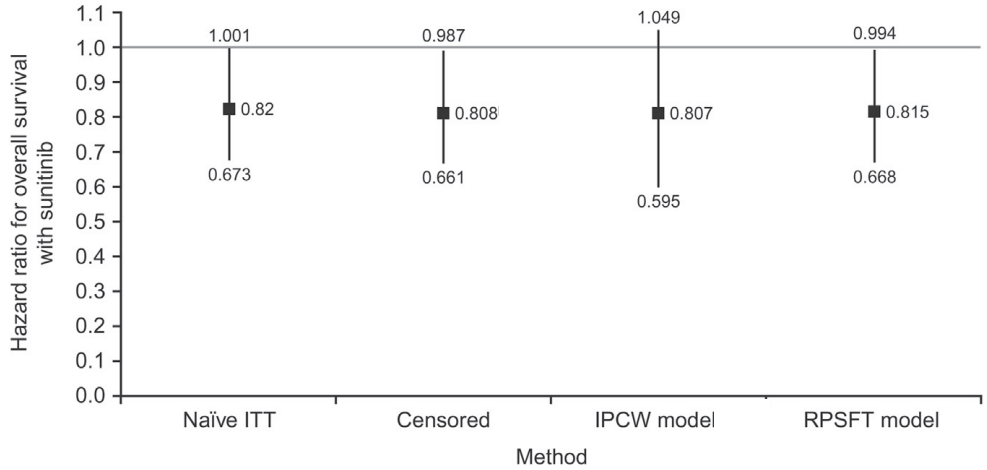


Figure 5: ITT and alternative analyses of the mRCC study

While the results are similar, it is notable that the IPCW model has wider confidence intervals as the censoring model and the impact of the RPSFT model was minimal, most likely because of the limited cross-over in the study.

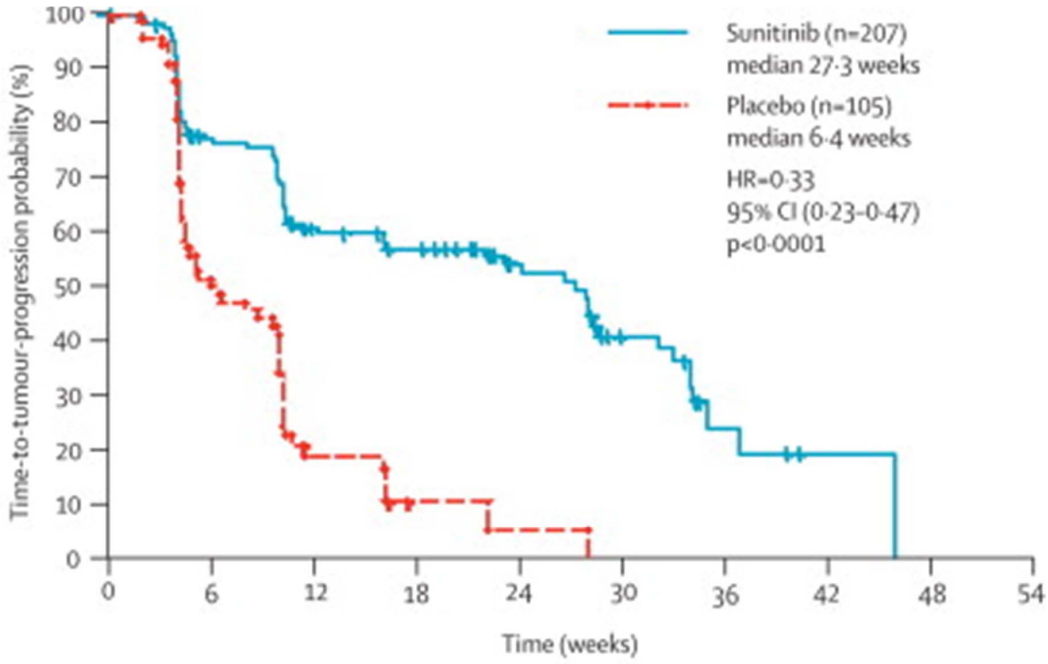
Sunitinib for gastrointestinal stromal tumors (GIST)

In this international phase III, multi-center, randomised, double-blind, placebo-controlled study of sunitinib for the management of gastrointestinal stromal tumors (GIST), 312 patients with with advanced and documented imatinib resistance were randomized to receive either sunitinib ($n = 207$) or placebo ($n = 105$).

Crossover was allowed after an interim analysis concluded a significant gain in the time to progression (Figure 6), all patients in the placebo arm were allowed to switch.

Sunitinib trial: Time to tumor progression

During the interim analysis, the following results were observed[9]



Number at risk

Sunitinib	207	106	67	53	34	18	5	1	0
Placebo	105	36	9	2	1	0	0	0	0

GD Demetri, AT van Oosterom, CR Garrett, et al. Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomised controlled trial. Lancet, **338**, 1329-1338. http://dx.doi.org/10.1016/S0140-6736(06)69446-4

Figure 6: Time to tumor progression at the interim analysis of the sunitinib trial for GIST management

[9] Demetri et al., Lancet 2006, Blay, Ann Onc 2010.

Sunitinib trial: Overall survival

The initial results of the study are shown in the following Figure[10].

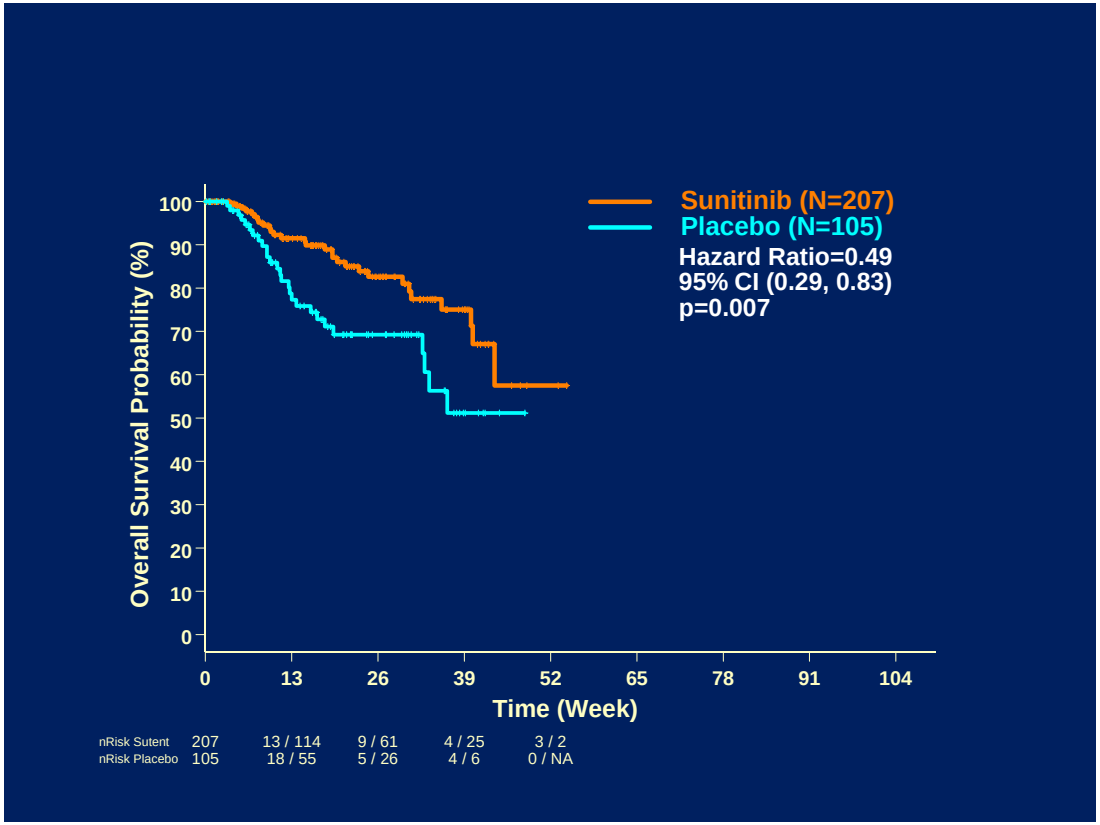


Figure 7: Overall survival in the sunitinib study (ASCO 2005)

[10] Huang & Xu, 2011.

Sunitinib study: Extended follow-up

Extended follow-up of the study patients was presented in the following year's ASCO conference as shown below:

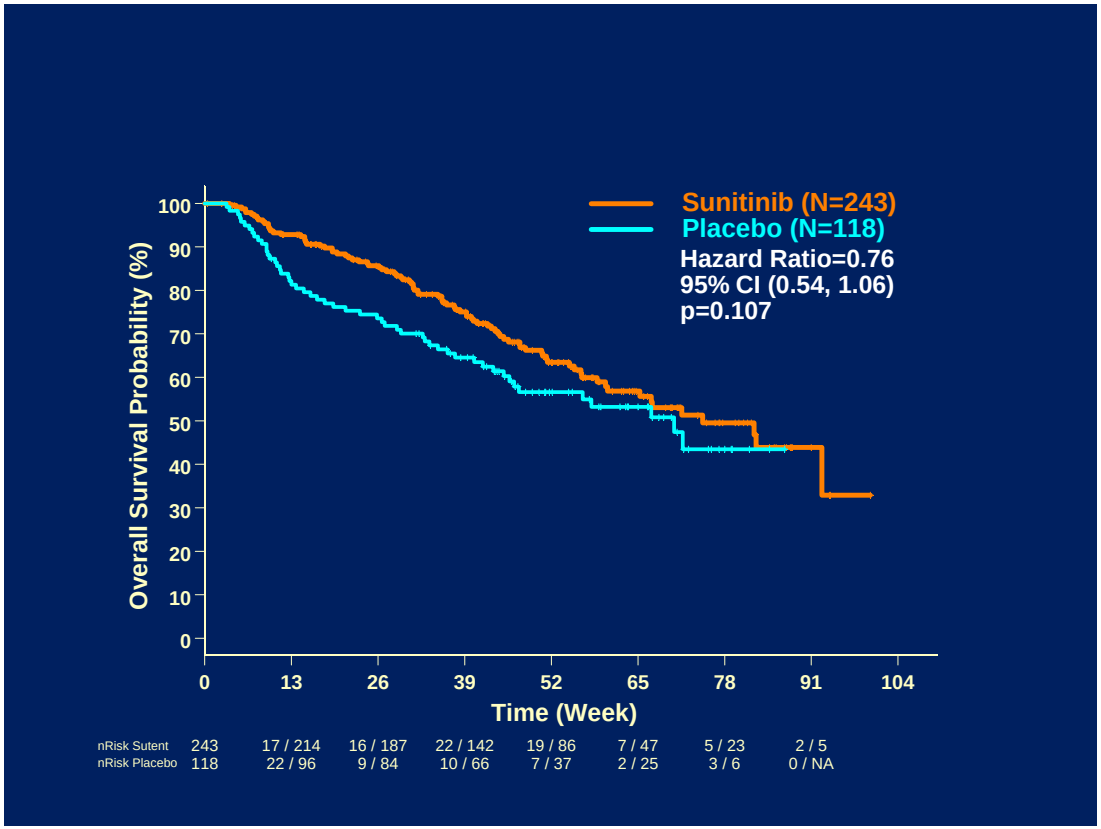


Figure 8: Overall survival in the sunitinib study (ASCO 2006)

Sunitinib study: Final results

... and again in 2008:

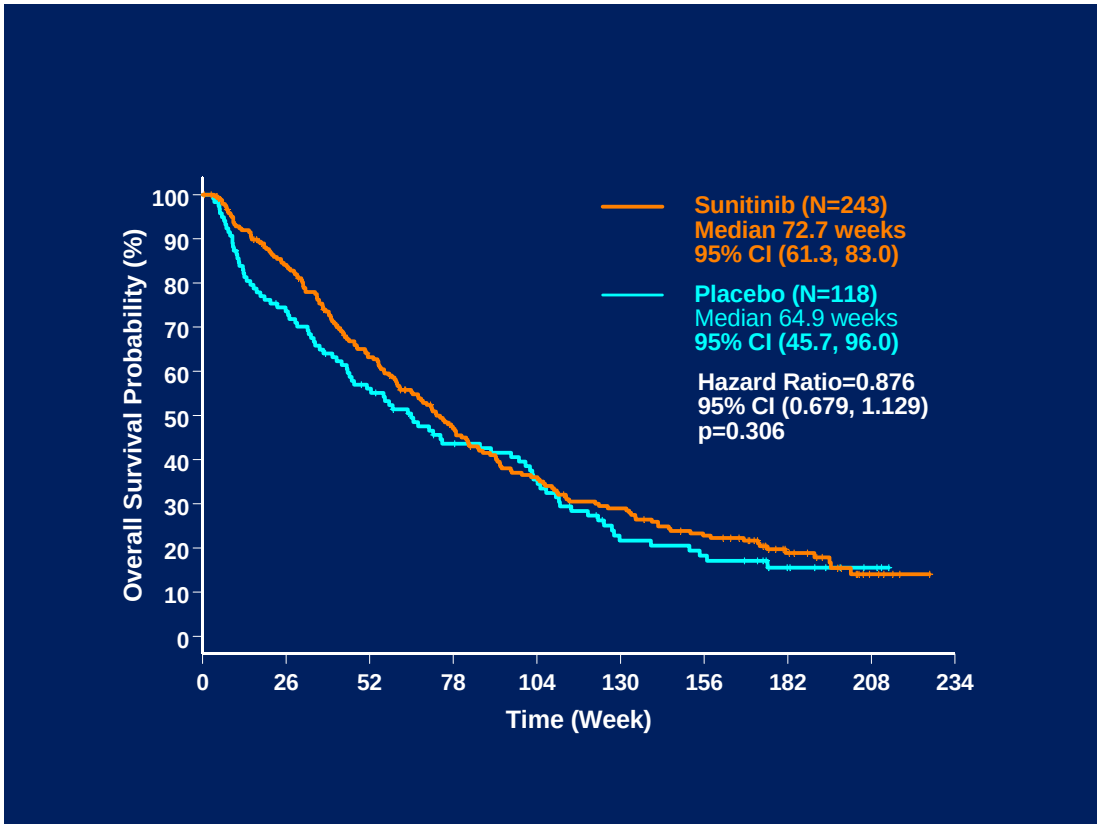
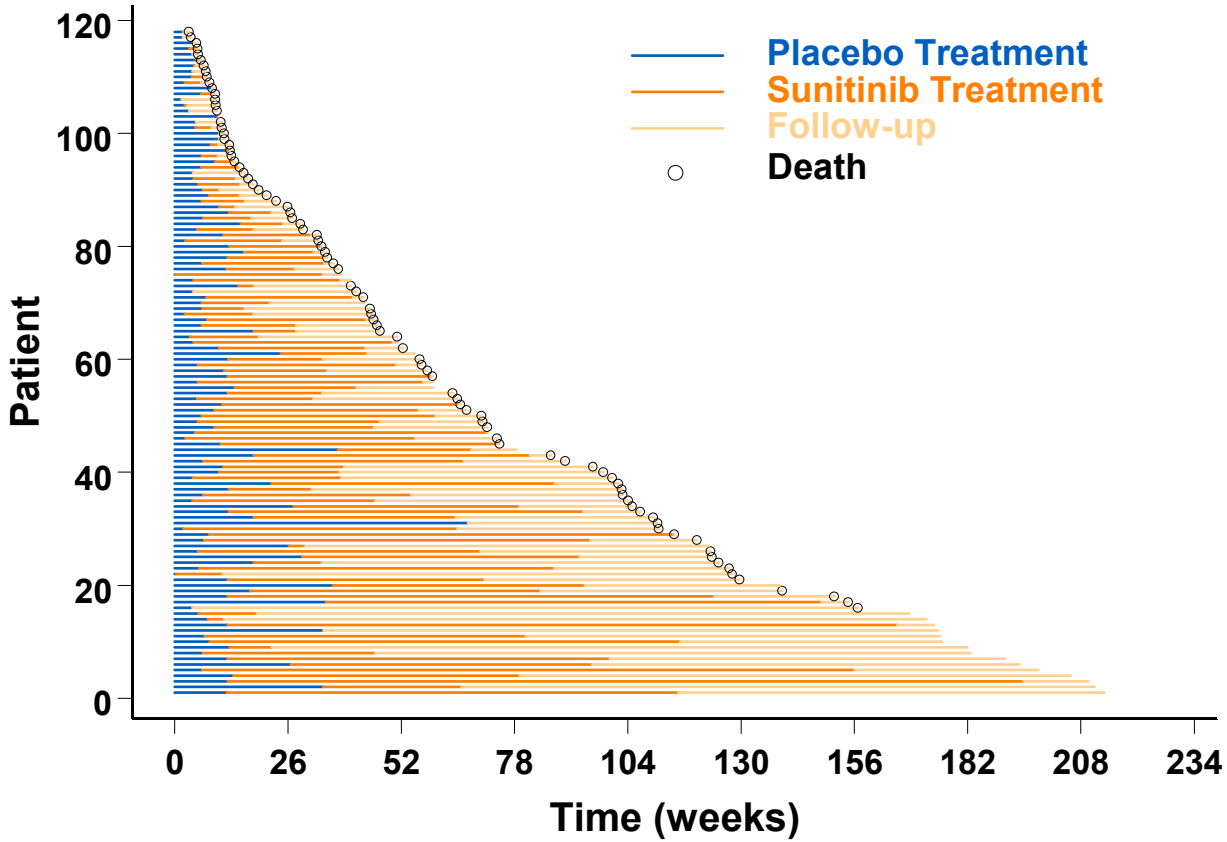


Figure 9: Overall survival in the sunitinib study: final results (ASCO 2008)

We can see a tremendous decay of the treatment difference as follow-up increases.

Sunitinib trial: the impact of switching

The extent and impact of switching from placebo to sunitinib is shown in Figure 10.



2nd Annual Pacific Coast Statisticians and Pharmacometricians Innovation Conference
July 15, 2011 Xin Huang & Qiang Xu

Figure 10: The extent and impact of switching

Concern over the impact of cross-over

The main concern of course is that the decay in the survival advantage was due to the high proportion of patients crossing over at progression (Figure 10).

Out of 73 patients in the placebo group whose disease progressed, 69 crossed over to sunitinib. In fact, the median time to progression in the non-sunitinib arm was less than two months (Figure 6)!

In an RPSFT model[11], the estimated value for $\psi = -0.656$ with the resulting hazard ratio was $\theta = 0.505$ (p=0.306).

This is to be compared with the hazard ratio of the usual ITT approach, where the hazard ratio of sunitinib versus placebo (which of course included many patients who switched) was $\theta_{ITT} = 0.876$ (p=0.306)

Note that the p-values in the ITT analysis and the RPSFT model are identical, even though the hazard ratio in the RPSFT model was substantially lower. This is by construction. The reason is the increased uncertainty induced by re-censoring.

[11] Demetri et al., 2012.

The Sunitimib study: RPSFT model

The result of the RPSFT model is as follows:

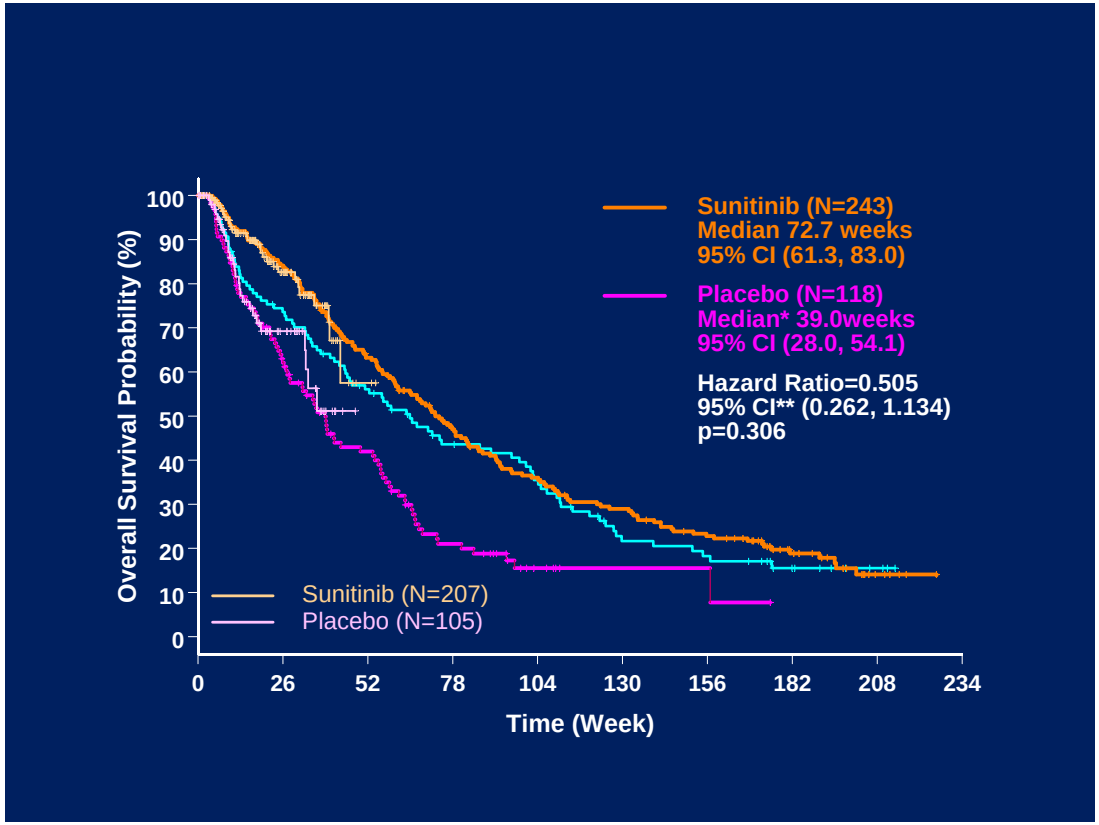


Figure 11: Adjusted results based on the RPSFT model.

Note how well the superimposed early survival curves fit the corrected survival curves. The RPSFT appears to have captured the early survival advantage of sunitinib.

References

1. Blay JY. Pharmacological management of gastrointestinal stromal tumours: an update on the role of sunitinib. *Ann Oncol* 2010, **21**:208–215.
2. Demetri GD, Garrett CR, Schöffski P, et al. Complete longitudinal analyses of the randomized, placebo-controlled, phase III trial of sunitinib in patients with gastrointestinal stromal tumor following imatinib failure. *Clin Cancer Res* 2012; **18**: 3170–3179
3. Huang X, Xu Q. Adjusting the crossover effect in overall survival analysis using a rank preserving structural failure time model: The case of sunitinib GIST trial. 2nd Annual Pacific Coast Statisticians and Pharmacometricians Innovation Conference. San Luis Obispo, CA, July 13-15 2011.
4. Jönsson L, Sandin R, Ekman M, et al.Comparative effectiveness research/health technology assessment (HTA). Analyzing overall survival in randomized controlled trials with crossover and implications for economic evaluation. *Value in Health* 2014; **17**: 707–713.
5. Motzer RJ, Hutson TE, Tomczak P, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med* 2007;**356**: 115–124.
6. Robins JM, Finkelstein DM. Correcting for non-compliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000; **56**:779–788.
7. Mark SD, Robins J. A method for the analysis of randomized trials with compliance information: An application to the multiple risk factor intervention trial. *Cont Clin Trials*; 1993; **14**: 79–97.
8. Sterne JAC, Tilling K. Gestimation of causal effects, allowing for timevarying confounding *The Stata Journal***2**: 164–182
9. Tsiatis A. ST 520: Statistical Principles of Clinical Trials. Lecture notes; chapter 8. Retrieved on 17 November, 2015 from www4.stat.ncsu.edu/d̃zhang/st520/520notes.pdf.
10. White I. Methods for handling treatment switching: rank-preserving structural nested failure time models, inverse-probability-of-censoring weighting, and marginal structural models. Presented at the HTMR network workshop on Methods for adjusting for treatment switches in late-stage cancer trials. London, 20th February 2012.