One-way optional crossover: biases and analysis approaches

Leen Slaets

Jan Bogaerts

EORTC

Overview

- The issue
- A famous example
- Framework, analysis methods, parameters at play
- A simulation study
- Concluding remarks

Progression Free Survival (PFS)
Overall Survival (OS)

- **“Approximate quotes” below:**
- **We should power the trial for PFS, because we cannot realistically expect to prove difference in OS. … The trial size would be too big.**
- **We now have several active follow-up treatments that will make interpretation of the OS data difficult.**
- **The sponsor can design the trial on PFS but should ensure sufficient power to detect OS effects.**

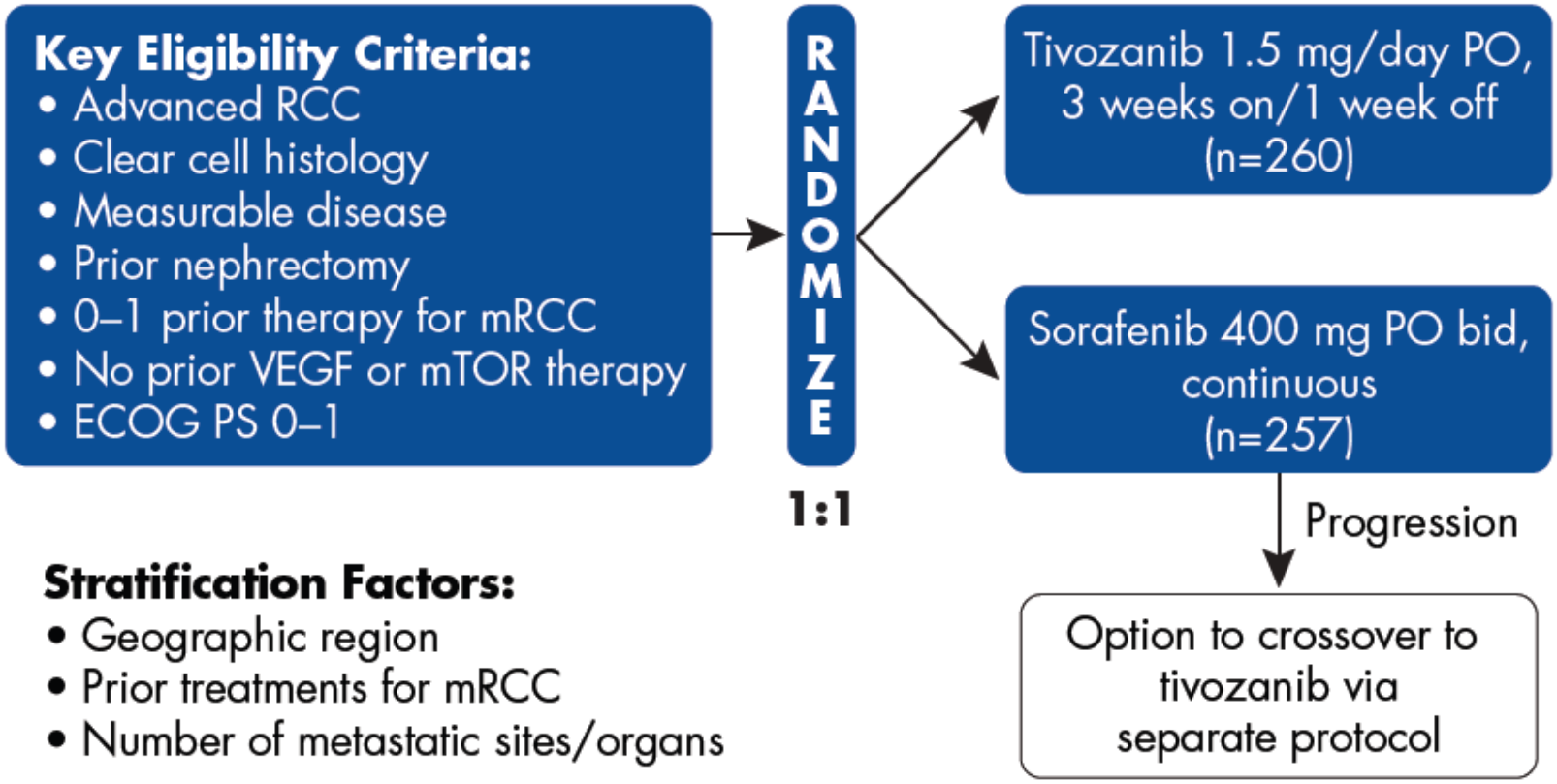
The issue

- **Broad: the influence of later lines of treatment on the OS comparison**
- **Narrow: one-way partial crossover at time of PD, from control arm to (class of) experimental drug**

- **Narrow is the subject of the following work**

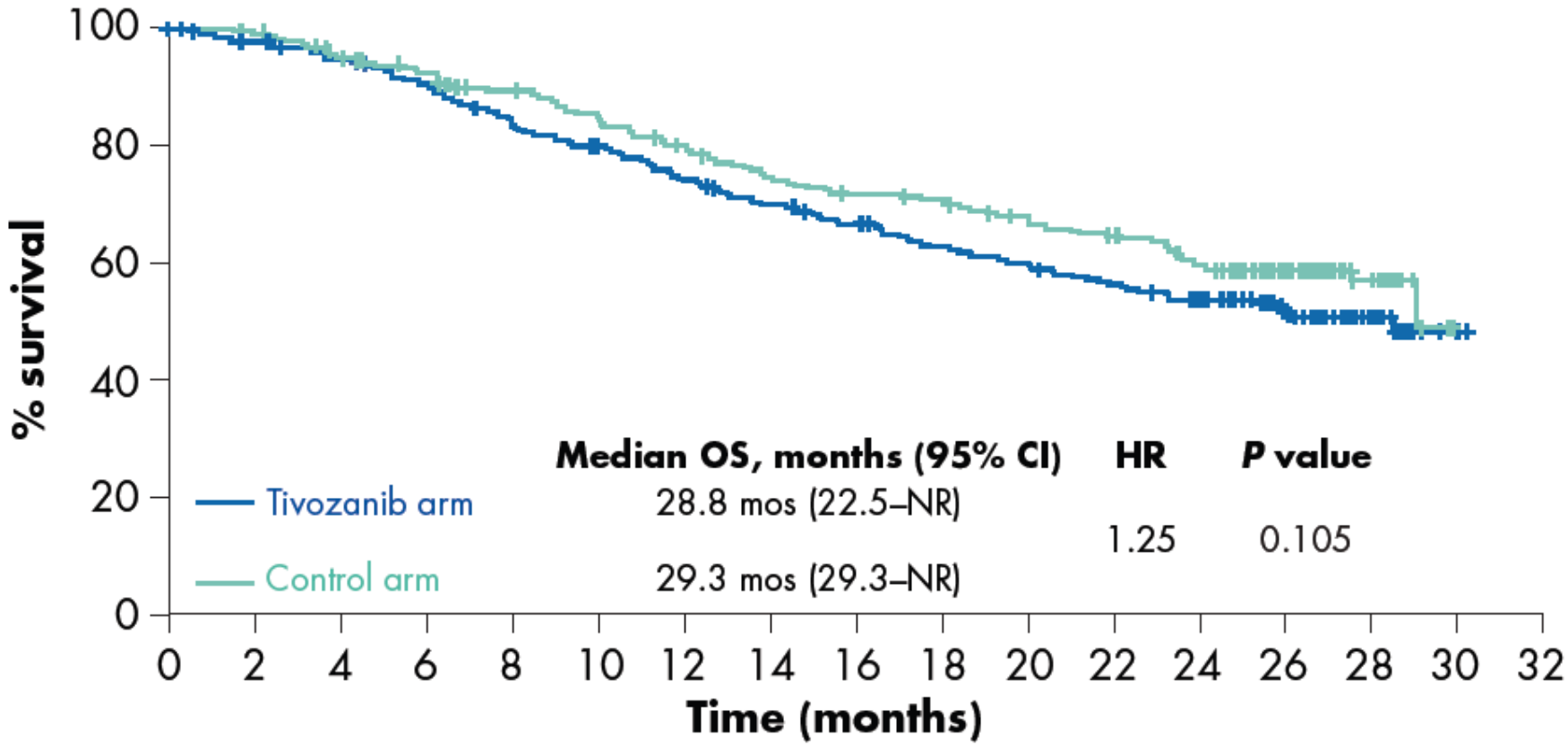
A famous example

Figure 1. TIVO-1: Phase III superiority study of tivozanib vs sorafenib as first-line targeted therapy for mRCC.



PFS HR = 0.80; 95% CI (0.64;0.99) P= 0.042, but …

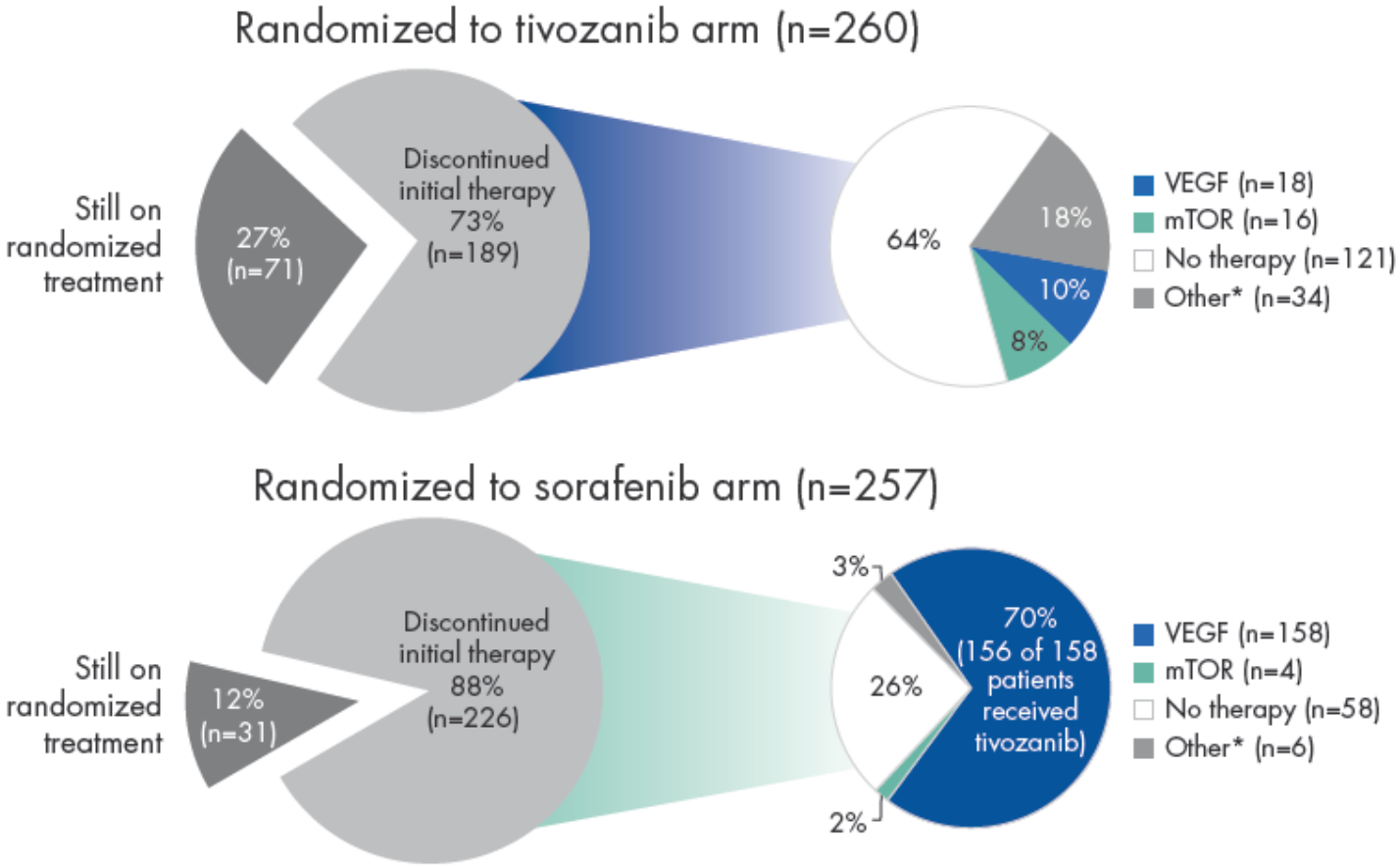
Figure 2. Protocol-specified, final OS analysis.



Patients at risk:

Tivozanib arm	260	256	241	227	211	198	183	170	159	148	142	133	125	89	39	2	0
Control arm	257	249	241	232	218	208	194	181	170	167	157	151	137	98	43	3	0

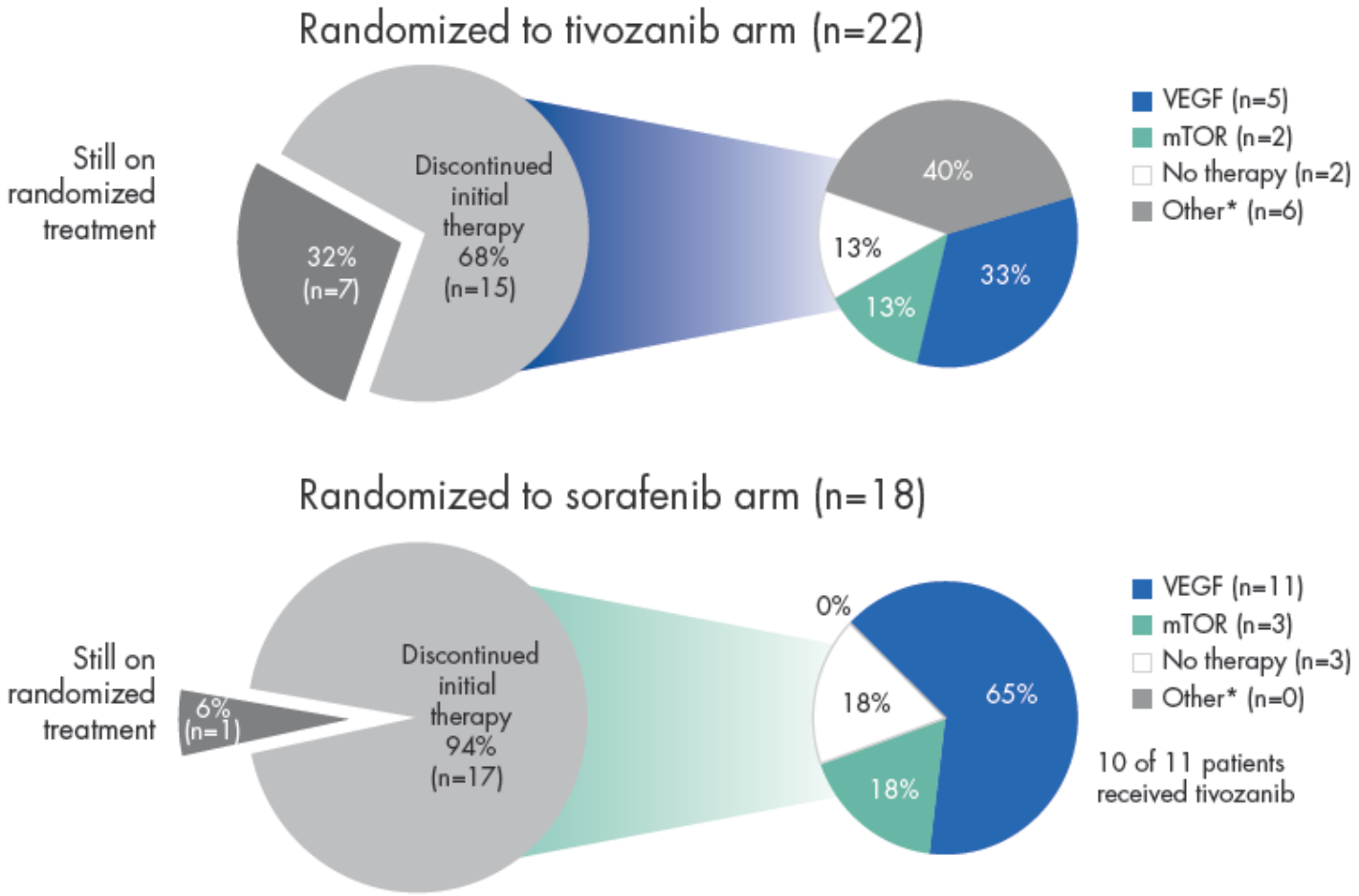
Figure 4. Next-line therapies.



[a]Other includes radiotherapy, cytokine, or other therapy.

Due to rounding, total does not equal 100%.

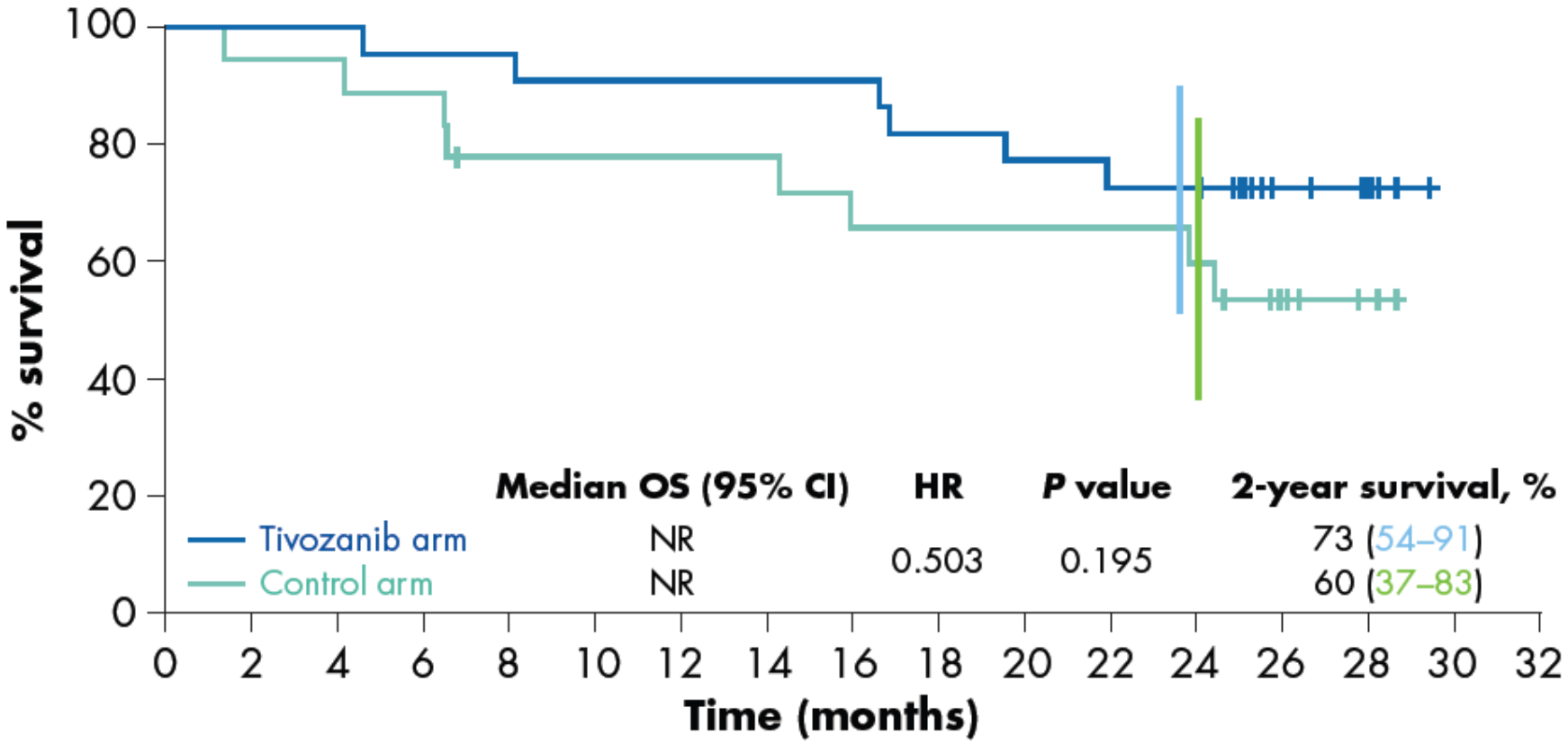
Figure 5. Next-line therapies in North America/Western Europe.



[a]Other includes radiotherapy, cytokine, or other therapy.

Due to rounding, total does not equal 100%.

Figure 3. OS in patients from North America/Western Europe.



Patients at risk:

Tivozanib arm	22	22	22	21	21	20	20	20	20	18	17	16	16	8	5	0
Control arm	18	17	17	16	13	13	13	13	11	11	11	11	10	6	3	0

AVEO Pharmaceuticals, Inc. (AVEO) - NasdaqGS

Add to Portfolio

2.10 Oct 30, 4:00PM EDT

Enter name(s) or symbol(s) | GET CHART | COMPARE | EVENTS | TECHNICAL INDICATORS | CHART SETTINGS | RESET

Basic Chart | Full Screen | Print | Share | **Send Feedback**

Elective cross-over at PD

- Some patients on the control arm get treated with the experimental drug (or other drugs from the same novel class) at the time of progression
- Most often this is explicitly made possible in the protocol
- But … the situation may be unavoidable, because of a class existing on the market
  - Question is whether this last situation needs correction

A paradox

- We undertake a trial to establish the potential value of a treatment (in comparative fashion).
- We design the trial so that we make it harder to prove, on the grounds that everybody (on the trial) needs to have access to the new treatment.
- The net result may be that only the participants on the trial get access to it.

- Question: who is “we”

Some simulation work

- **Purpose: compare and qualify methods used to answer the counterfactual question: "What would the OS effect have been in absence of cross-over ?"**
- **AKA: putting the toothpaste back in the tube**

Simulations: framework

- **It is a missing data problem:**

  We will never observe the outcome of interest:

  OS in absence of crossover

- **Popular presumption:**

  due to the crossover, the treatment effect observed on OS in the trial is much smaller compared to the “true” unobserved treatment effect (in absence of crossover)

  Especially when a strong DFS benefit did not result in a OS benefit, crossover is “scapegoated”.

  And a variety of sensitivity analyses are reported.

Simulations: Goal

By means of the observed trial data and simulation of unknown parameters we try to gain insight in the following:

- Is it realistic to assume that the lack of OS effect in the ITT analysis is completely due to the crossover?

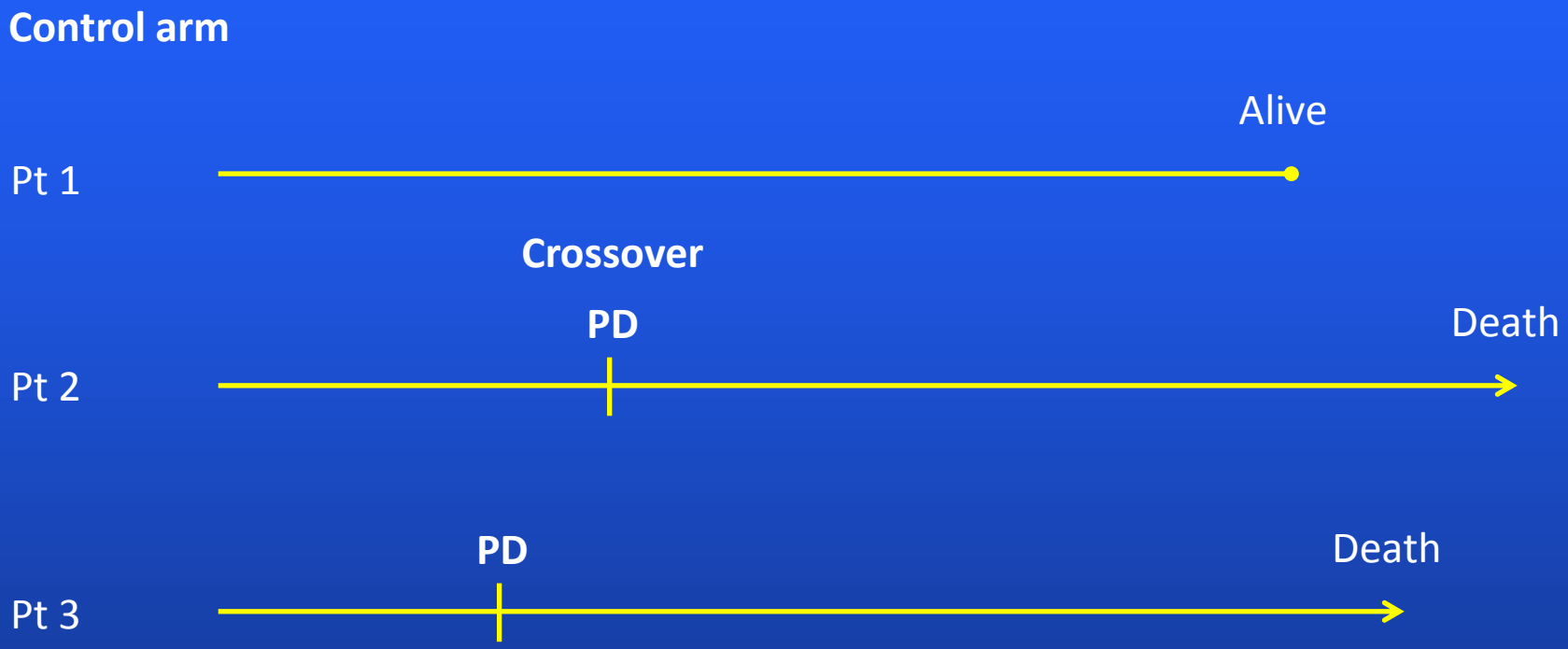
  What if the next-line standard therapy is more effective in the control arm?

  (e.g. due to resistance mechanisms triggered by the experimental treatment)

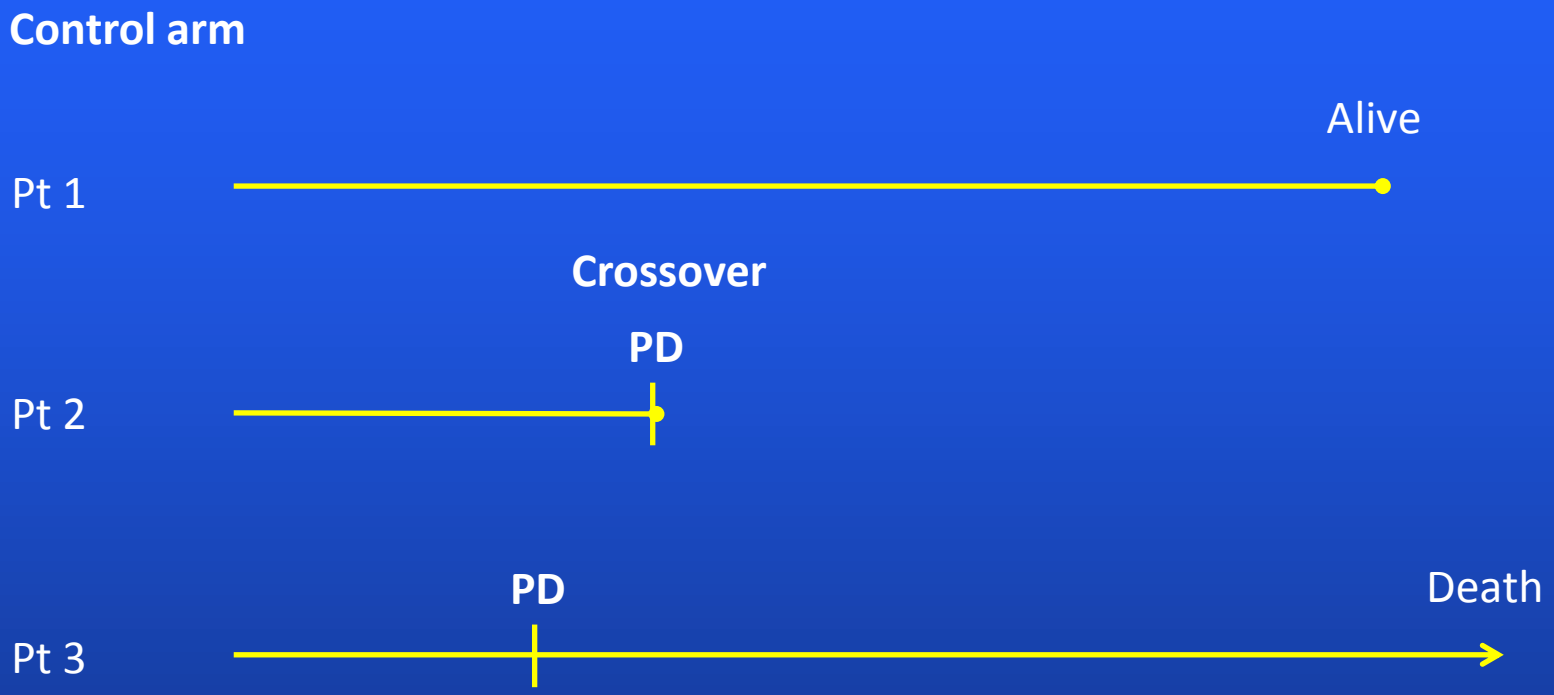
- What can we expect from popular sensitivity analyses.

  How do they behave under various (untestable) assumptions.

Methods: ITT (as is)

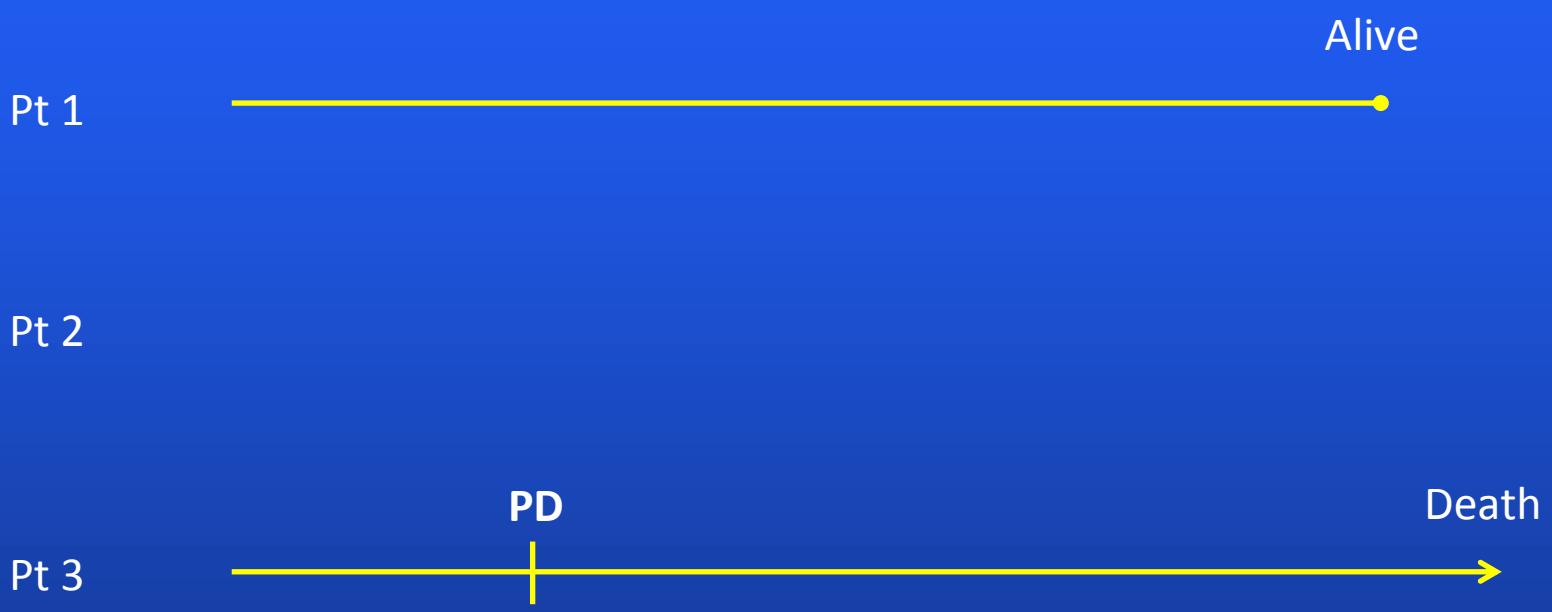


Methods: censor at crossover

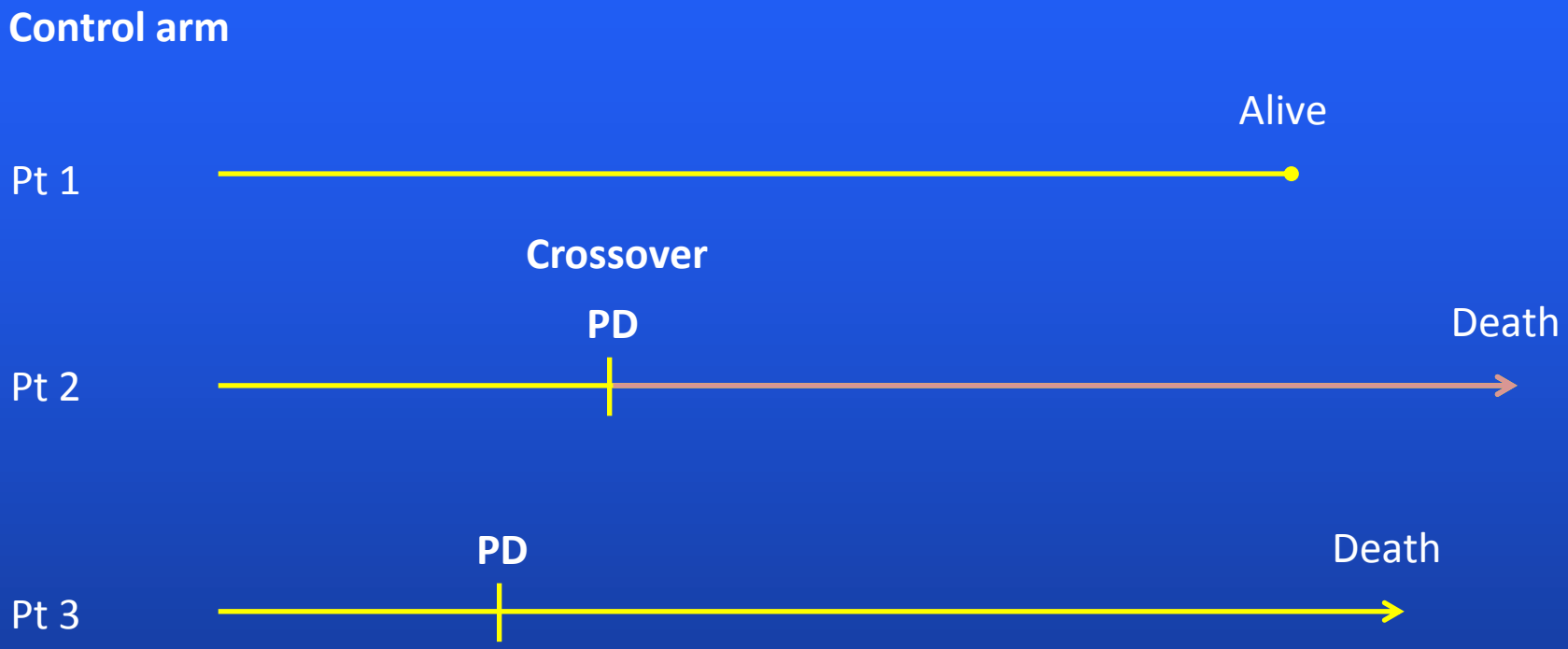


Methods: exclude

Control arm



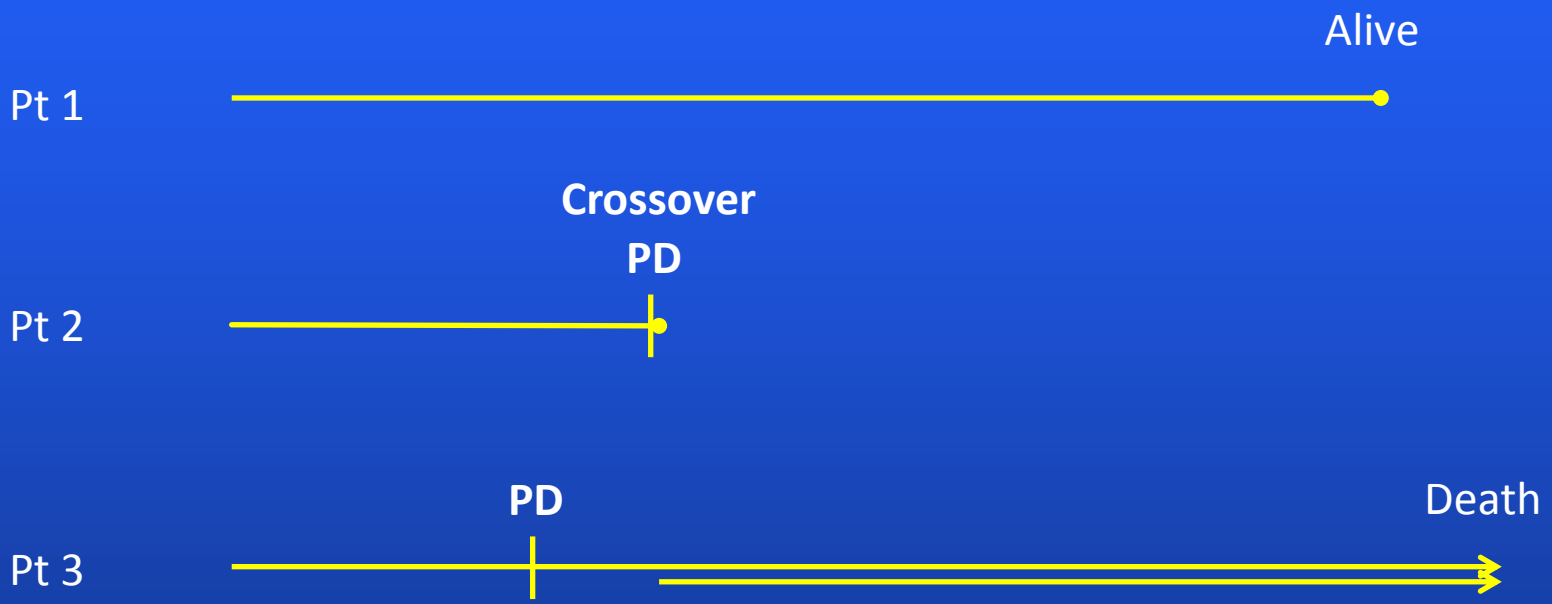
Methods: time dependent covariate



Red = experimental arm

Methods: IPW

Control arm



We attribute more weight to similar patients who did not cross.

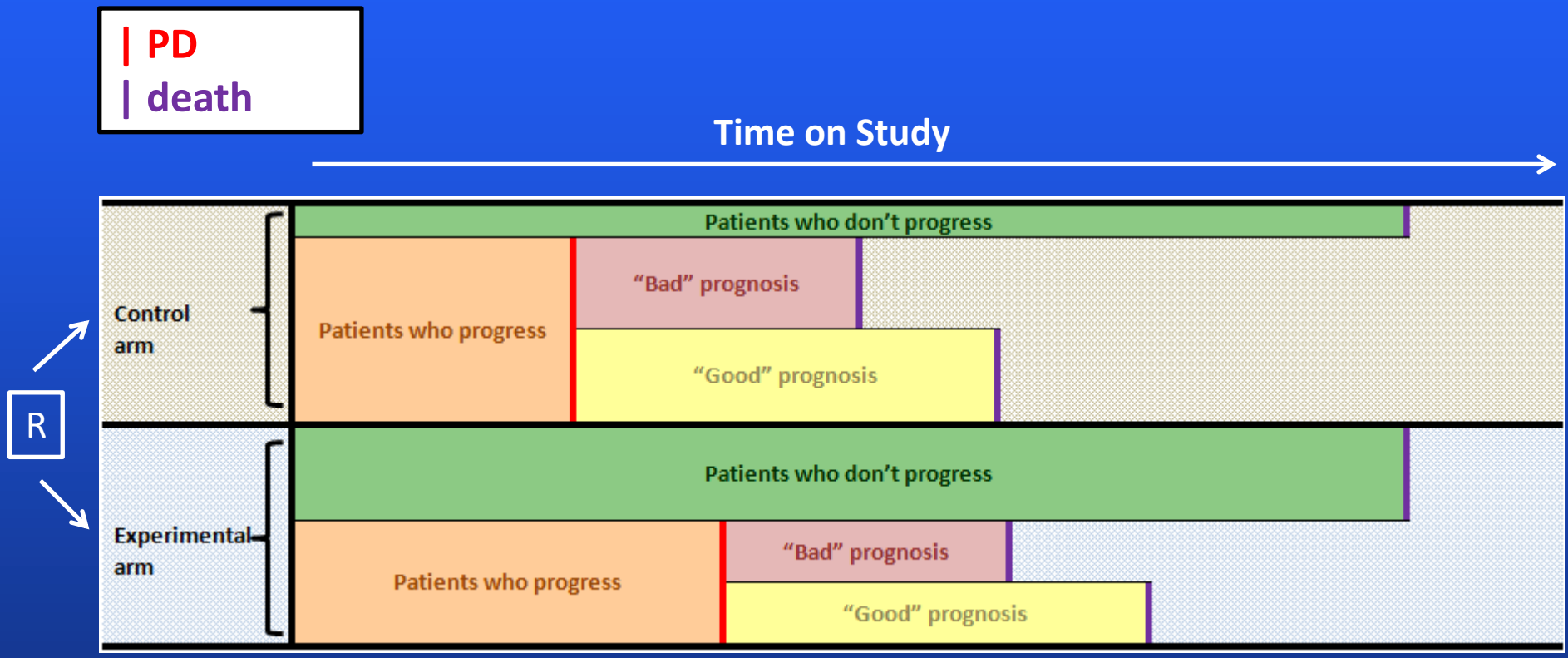
this is schematic!

Simulations: some terminology

- **Next-line** = the therapy given at PD
- **Next-line standard therapy**= the current standard therapy(s) for patients who have PD.

Simulations: schematic model

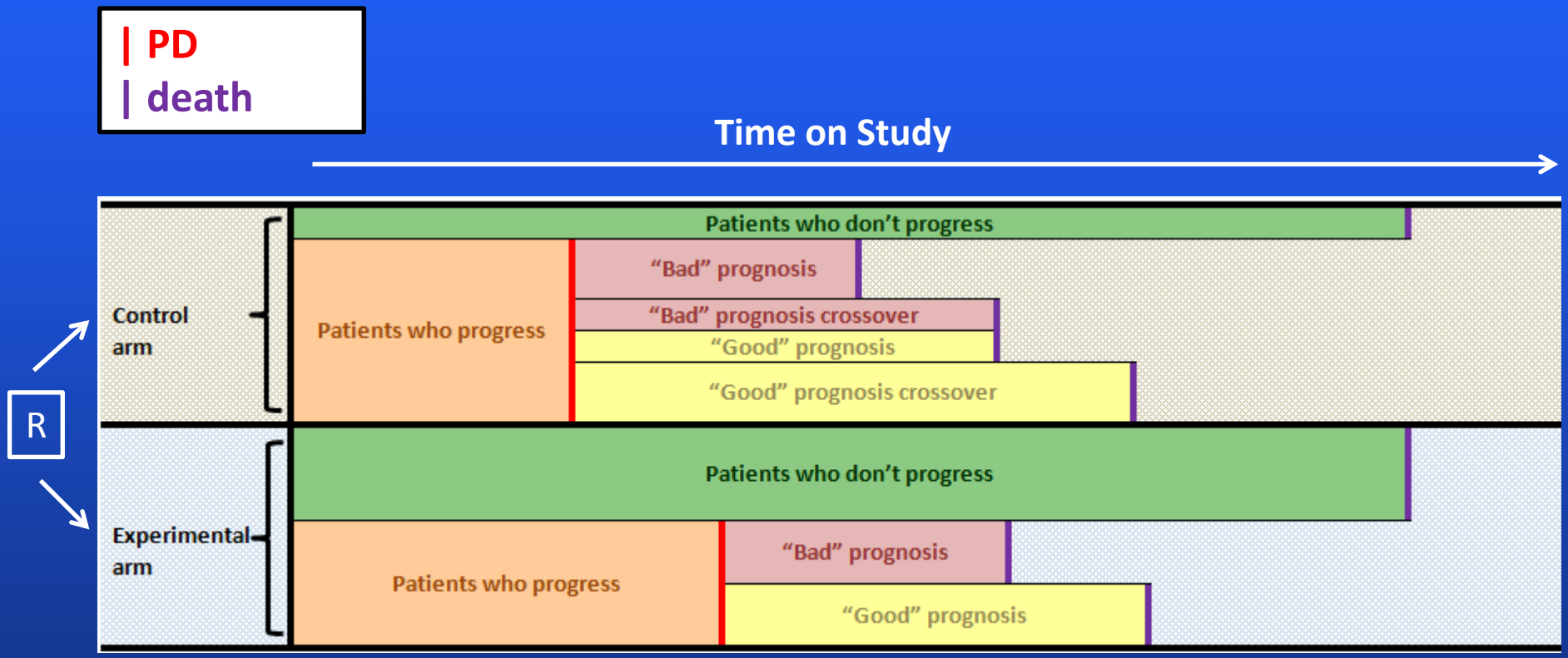
No Crossover



This is a simplified schematic presentation!

Simulations: schematic model

Crossover



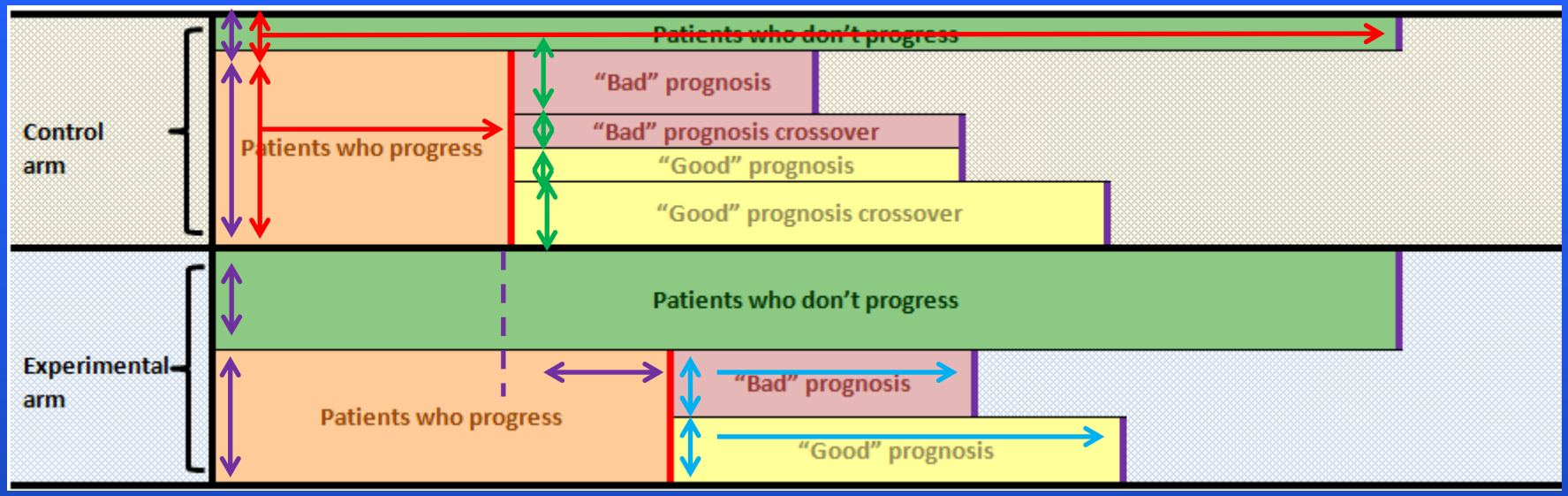
This is a simplified schematic presentation!

Simulations: setting & parameters

- **All simulation parameters can be classified in 3 types:**
  - Some are known (observed in the trial) and can be fixed.
  - Others can be given a reasonable estimate
  - Others are unknown: these are important investigational parameters

- **All these parameters nees to be provided in an R function. This function then performs a customized, automated simulation study**

Simulations: observed parameters

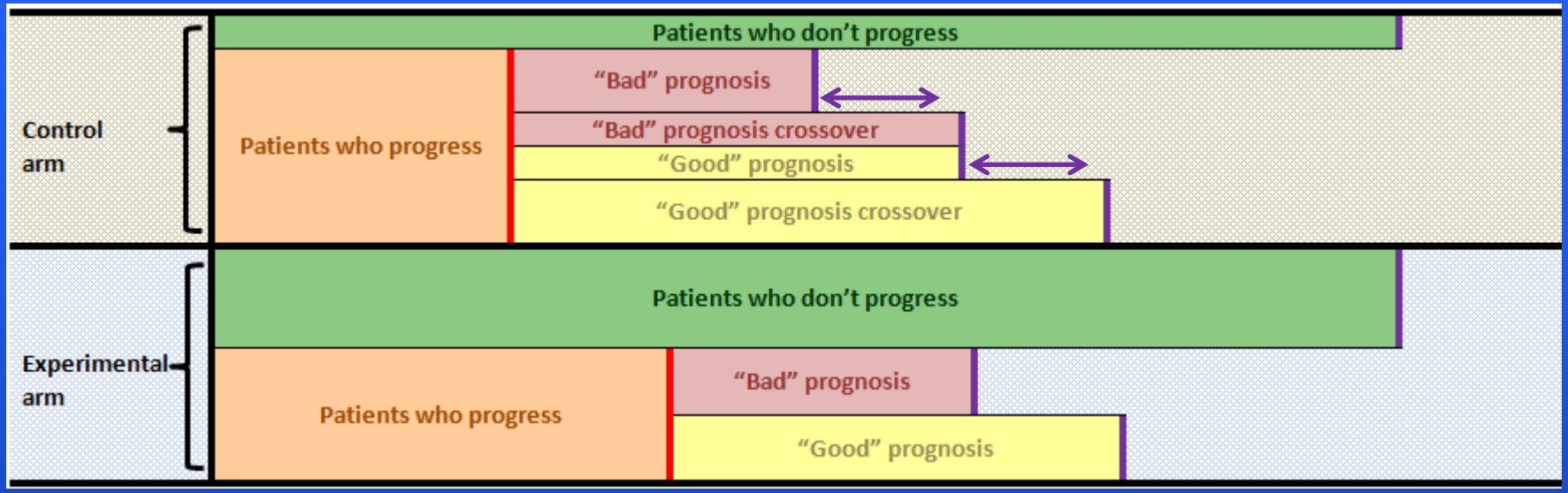


- Number of patients
- OS analysis timing (in terms of number of events)
- Accrual period duration

} design related

- Treatment effect on PFS exp. vs control (Hazard Ratio)
- X year PFS rate, control arm
- X year post PD OS rate, experimental arm
- % crossover at PD (in control arm)

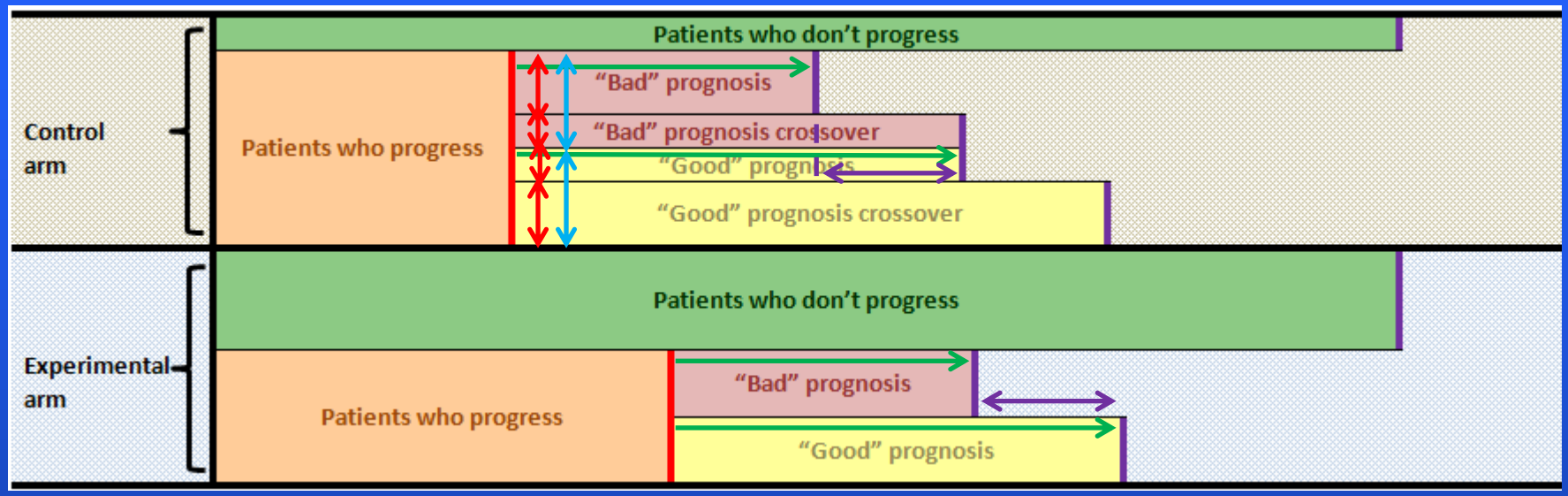
Simulations: estimable parameters



- Exp. vs. standard therapy as next-line in control arm (Hazard ratio)
  We assume this is the same for "bad" and "good" prognosis patients

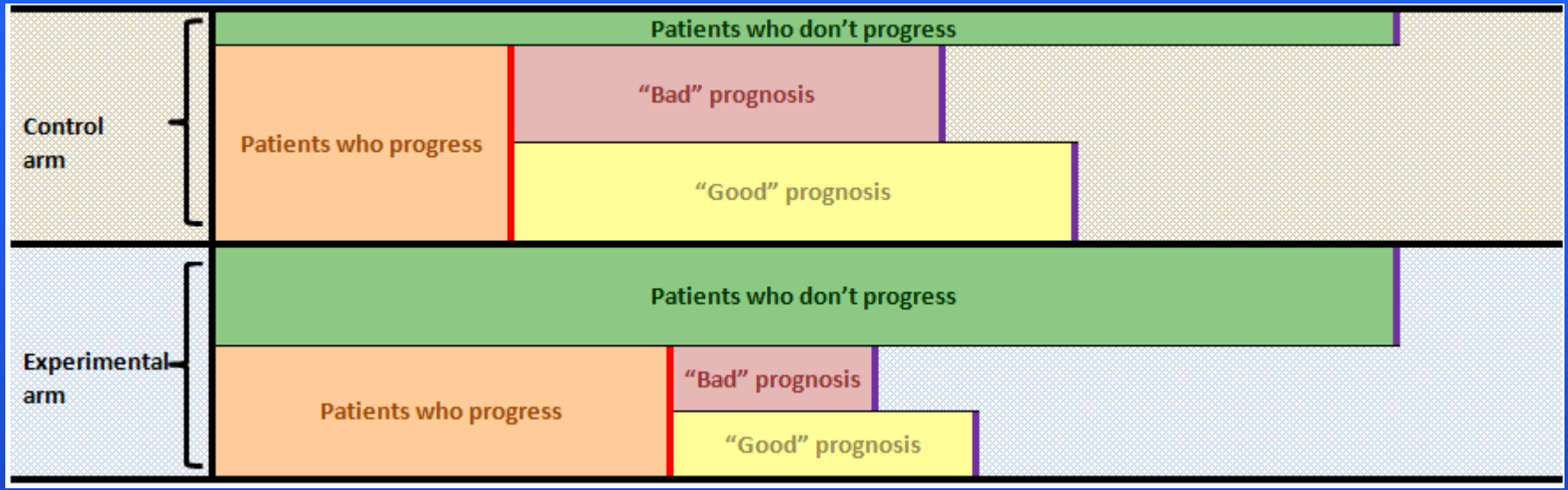
  "Crossover effect"

Simulations: unknown parameters



- prognostic effect on OS post PD, "good" vs. "bad" prognosis patients (Hazard ratio)
- Relative risk for crossover "good" vs. "bad" prognosis patients (RR=2 on this slide)
- Relative amount of "good" vs. "bad" prognosis patients
- Next line standard therapy effect on OS, exp. vs control (Hazard Ratio)

Simulations: unknown parameters



This slide: next-line standard therapy is less effective in exp. (vs. control) (HR > 1)

Previous slide: next-line standard therapy is equally effective in exp (vs. control) (HR = 1)

Simulations: Analysis methods:
how are they affected by the unknown parameters?

Unknown parameter	ITT	Sensitivity analyses (cens., IPW,…)	True (no cross)	
Next-line standard therapy effect (exp. vs control)	X	X	X	**"overlooked"**
Exp. vs. standard therapy as next-line in control arm ("crossover effect")	X	X*		**"scapegoated"**
Relative risk for crossover ("good" vs "bad prognosis")		X		**"Can of worms"** How selective the crossover is and in what direction
Strength of prognostic effect on OS ("good" vs "bad prognosis")		X		
Relative amount of "good" vs "bad prognosis" patients		X		

* Only for the method: treatment as a time dependent

Simulations: Example 1 (multiple settings)

- **Treatment effect on PFS (exp.vs control) : HR=0.7**
- **Next line standard therapy effect on OS (exp. vs control): HR=1**
  (next-line standard therapy equally effective in control and experimental)
- **50% crossover at progression**
- **"crossover effect" (exp. vs. standard therapy as next-line):**
  HR = 0.6, 0.8, 1, 1.25, 1.67
- **Relative risk of crossover at PD ("good" vs "bad" prognosis): RR=4**
  RR = 0.167, 0.25, 1, 4, 6
- **Strength of prognostic effect on OS ("good" vs "bad" prognosis):**
  HR = 0.10, 0.3, 0.5, 0.7
- **Other parameters: chosen to mimic an adjuvant setting in advanced breast cancer.**

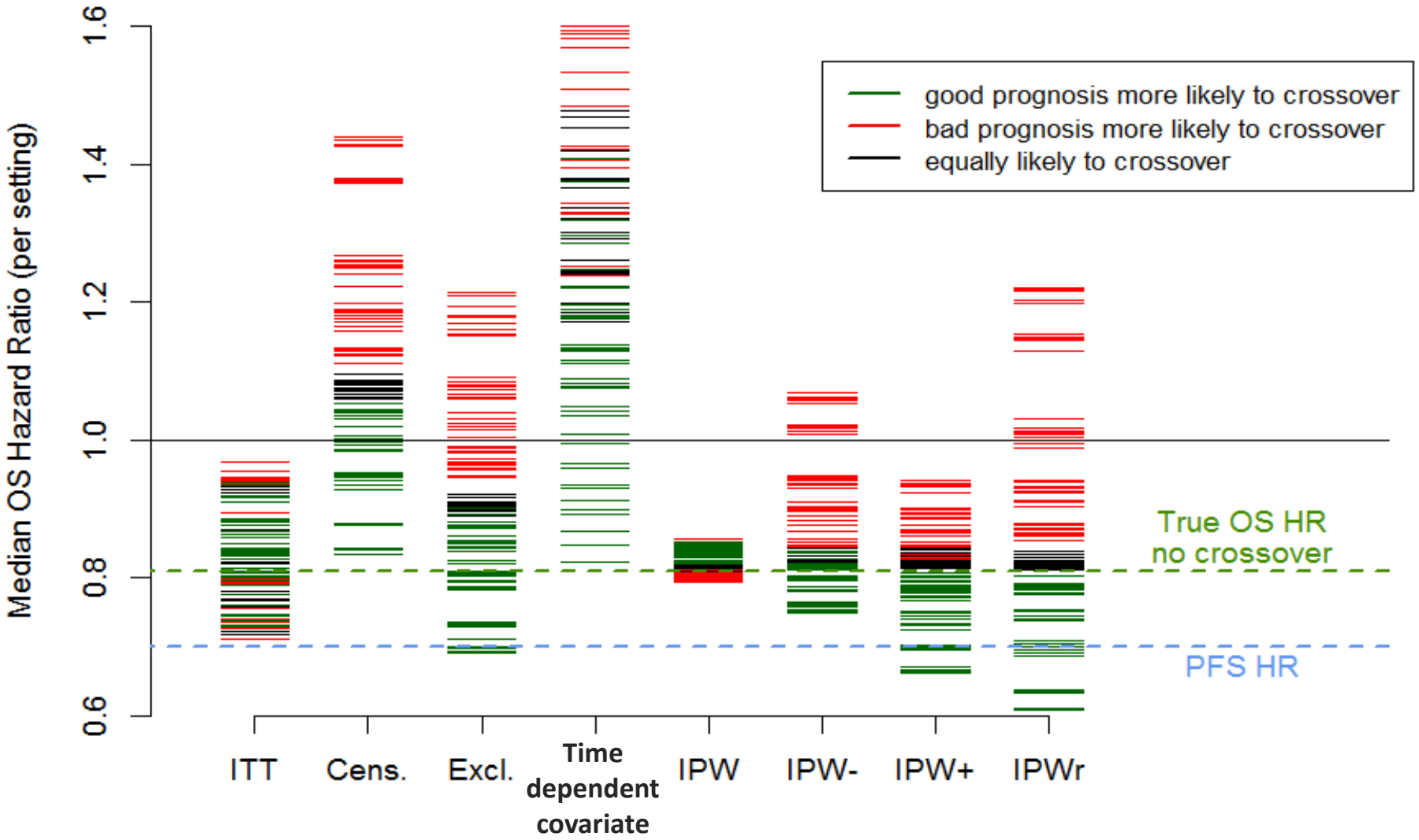
Simulations: Example 1 (multiple settings)

Note that for this simulation study with multiple settings:
we fix the next line standard therapy effect on OS (HR=1).

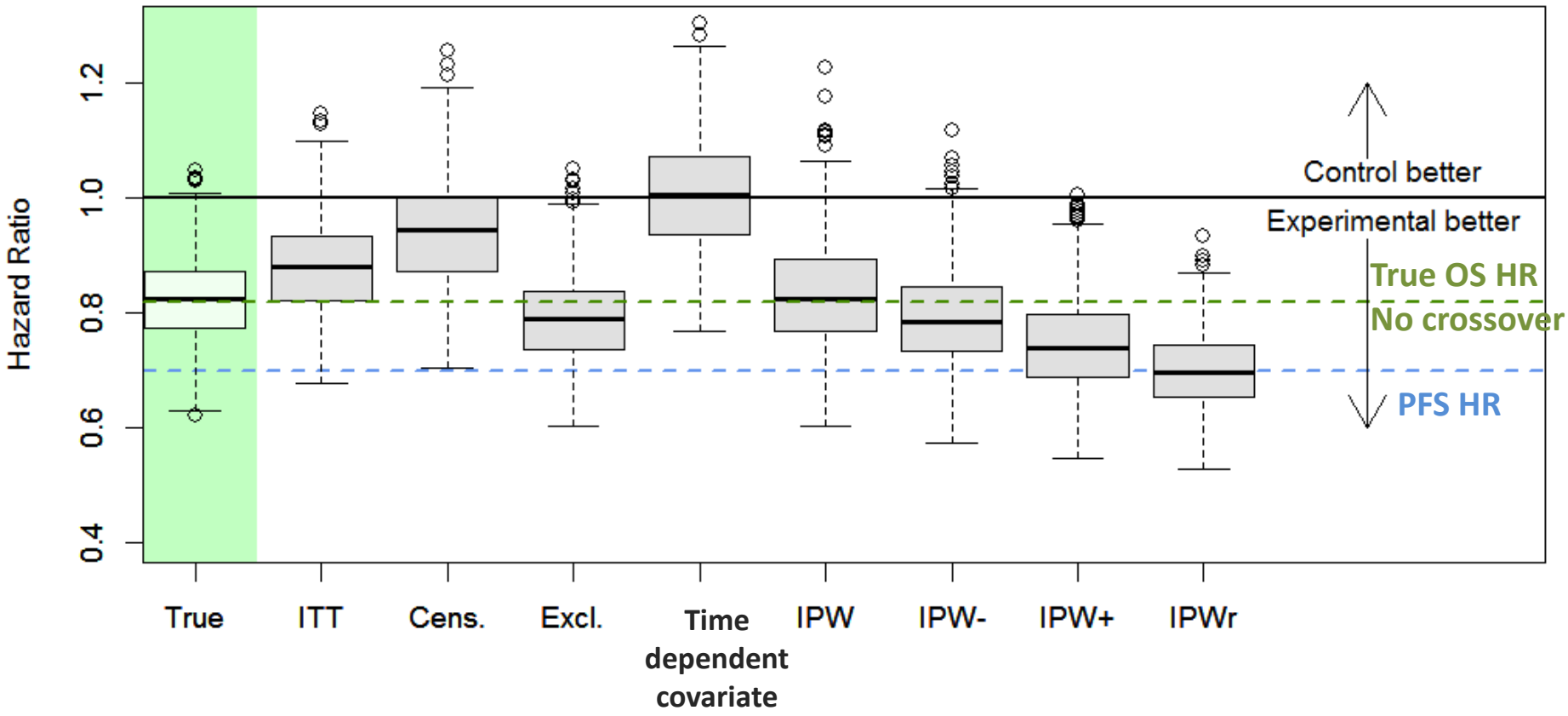
This parameter is also unknown, but has nothing to do with crossover.

However, it affects the true OS effect (also in absence of crossover)

Simulations: Example 1, multiple settings



Simulations: Example 1, 1 setting



Current learning from simulation(s)

- **Excluding patients, censoring at crossover, time dependent arm are bad methods for the counterfactual question**
  - How bad they are depends on the settings
  - … we tried lots of settings, quite broadly
  - Note that even when the crossover is not selective (black lines), these analyses are biased:

    these methods exclude/censor patients that had a PD (=selective)
- **In fact ITT performs fairly well**

- **IPW with full information works very well … but we are extremely unlikely to have the full information**

  If "good" prognosis patients are more likely to crossover, IPW is biased in favor of experimental.

  Bad IPW attempts are "rewarded" with a bigger bias in favor of experimental

Current learning from simulation(s)

- **Under the assumption that "good" prognosis patients are more likely to crossover and that crossover patients benefit from experimental (experimental better than standard as next-line):**

  A window between ITT and IPW gives a broad indication of where the true value is likely to be

  This holds regardless of whether next line standard therapy is less or more effective in the experimental vs. control arm.

Simulations: Example 2, Tivo-1

Observed trial results:

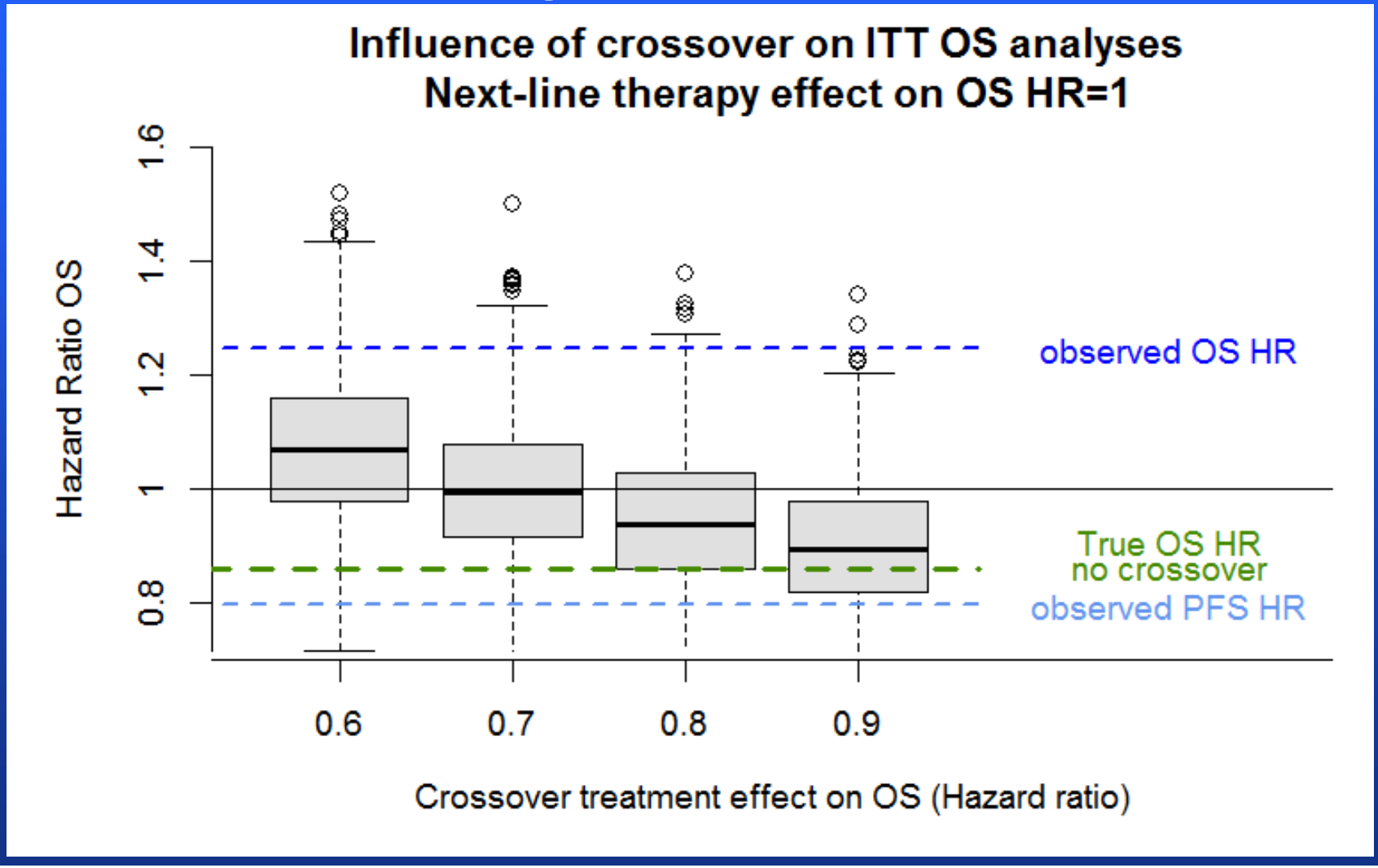
- **Treatment PFS comparison (ITT): HR=0.797**
- **Treatment OS comparison (ITT): HR=1.245**
  (OS better in control arm)

70% of patients with PD in control arm, crossed over to experimental treatment

To which extend can the observed OS effect (ITT) be explained by crossover?

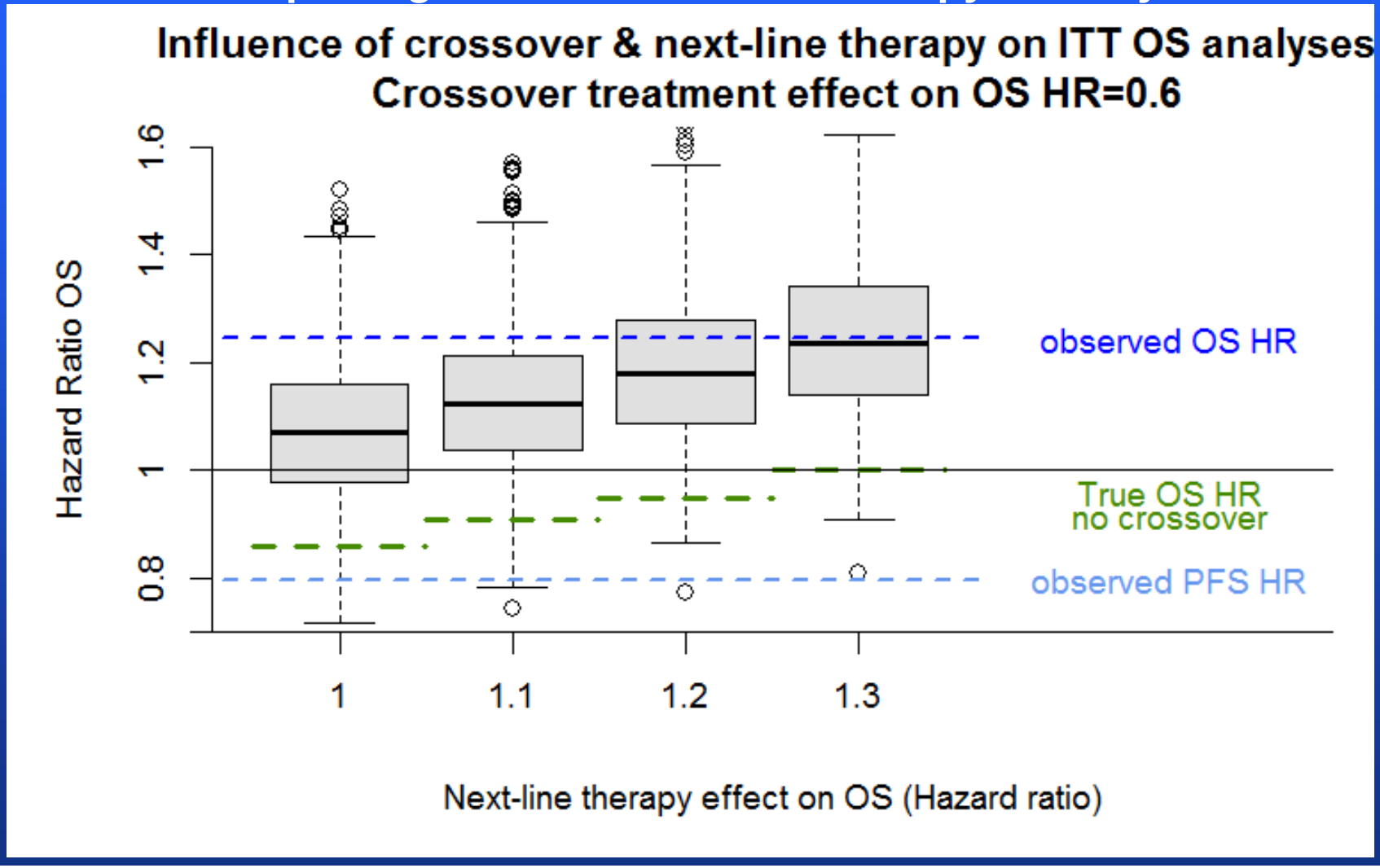
Simulations: Example 2, Tivo-1 ITT

Assuming next-line standard therapy equally effective in control and exp.
Exploring the "crossover effect"



Simulations: Example 2, Tivo-1 ITT

Assuming crossover treatment effect hazard ratio=0.6
Exploring next-line standard therapy efficacy



Simulations: Example 2, Tivo-1 ITT

- Based on the available Tivo-1 study data we could not simulate the observed OS effect when only exploring the "crossover effect"

- A combination of a "crossover effect" in the control arm and reduced next-line therapy efficacy in the experimental arm vs control yielded results similar to what was observed.

- For the latter setting, the "true" simulated OS effect (in absence of crossover) was: HR = 1.

Some more comments

- When to address the issue?
- Afterwards:
  - Far from evident
  - Sensitivity analysis to ITT as per our simulations
  - However: the follow-up treatment confounding remains a valid question
- Upfront:
  - Not allow crossover
  - “Regulate” or address follow-up treatment
  - See eg. AVEO: it was suboptimal to run the trial mainly in eastern EU

Conclusions

- To non-statisticians the issue seems to look ‘not hard for stats’
- Create more awareness
- Societal dimension to access to treatment

  < > statistical / methodological demands
- As a statistician: be aware of the issue, from design stage onwards

References

- Rimawi M. and Hilsenbeck S.G. (2011). J Clin Oncol. Feb 1;30(4):453-8
  Making Sense of Clinical Trial Data: Is Inverse Probability of Censoring Weighted Analysis the Answer to Crossover Bias?

- Robins J.M. and Finkelstein D.M. (2000). Biometrics 56, p.779-788.
  Correcting for Non-compliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring weighted (IPCW) logrank tests

- T. Fojo, L. Amiri-Kordestani, S. E. Bates (2011). JNCI vol. 103 (23), p. 1738-1740.
  Potential Pitfalls of Crossover and Thoughts on Iniparib in Trile-Negetive Breast Cancer.

- R. J. Motzer, D. Nosov et al. (2013). J Clin Oncol. (Published ahead of print)
  Tivozanib versus Sorafenib as Initial Targeted Therapy for Patients With Metastatic Renal Cell Carcinoma: Results from a Phase III Trial.

- R. J. Motzer, T. Eisen et al. (2013). ASCO GU, poster no 350.
  Overall Survival from a Phase III study of Tivozanib Hydrochloride versus Sorafenib for Metastatic Renal Cell Carcinoma.