Methods to estimate survival time after treatment switching in oncology – overview and practical considerations

Norbert Hollaender

Novartis Pharma AG, Oncology Global Development, CH-Basel
(email: norbert.hollaender@novartis.com)

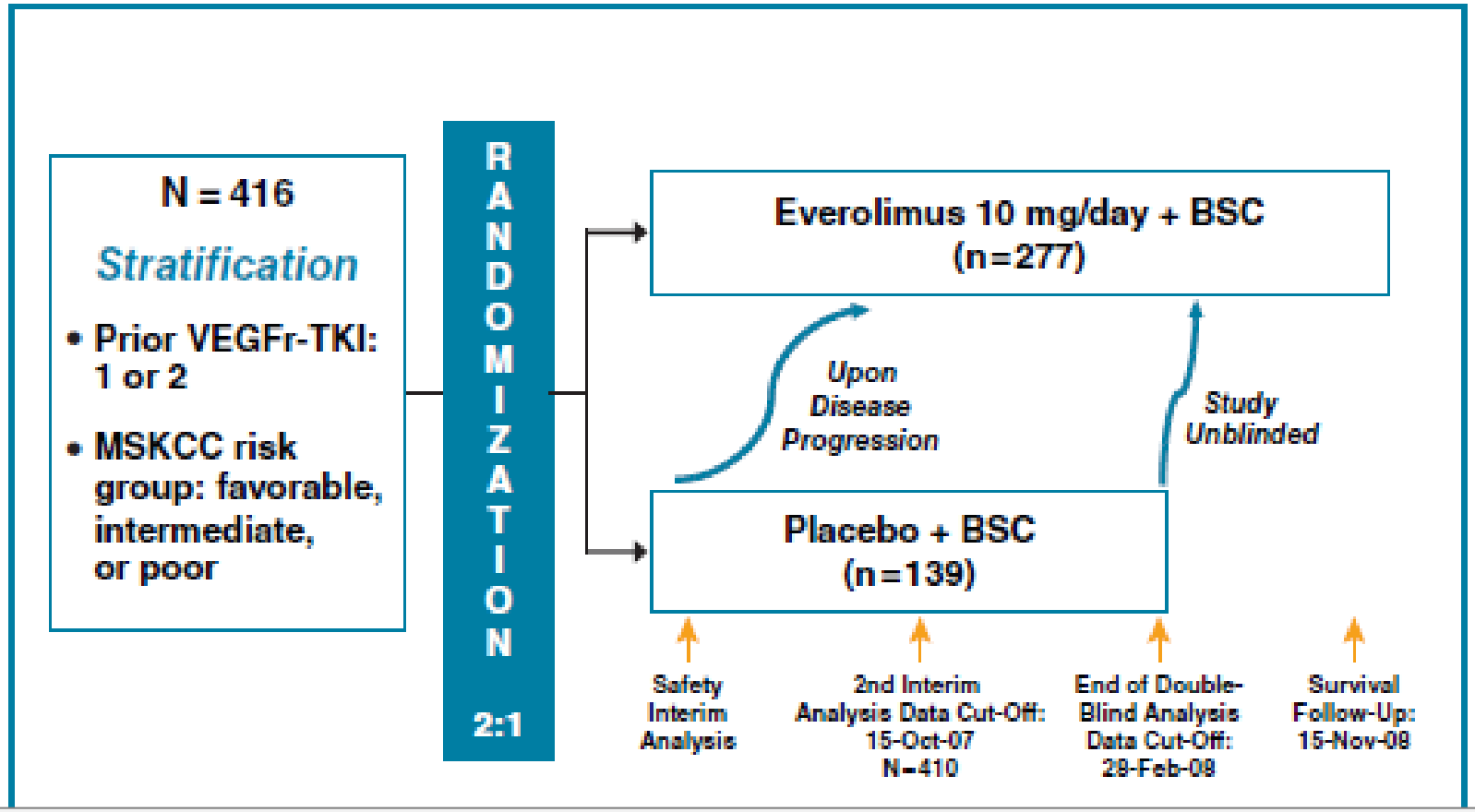
IQWiG in dialogue, Köln, 27-Jun-2014

NOVARTIS

Outline

- Motivation (RECORD-1 study)
- Simple (‘naive’) methods
  - Intent to treat analysis
  - Exclude treatment switchers
  - Censor switches at time of ‘cross-over’
  - Time varying treatment variable
- Complex methods
  - Inverse-probability-of-censoring weighting (IPCW)
  - Rank Preserving Structural Failure Time (RPSFT) Model
- Summary & Points to consider
- References

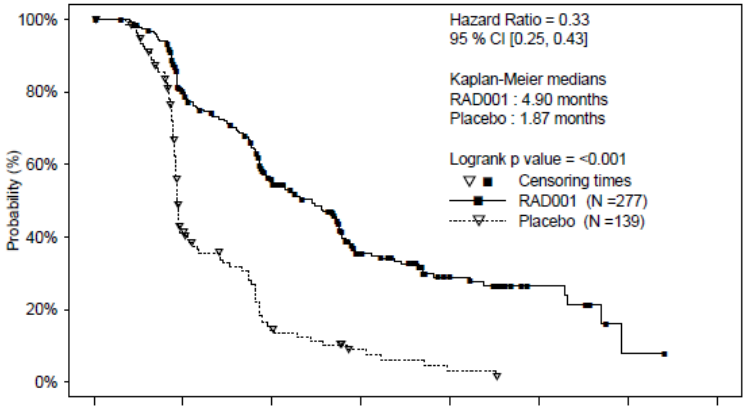
Motivation: RECORD-1 study comparing everolimus vs. placebo in metastatic kidney cancer



RECORD-1 study design allows to switch from Placebo to everolimus (RAD001) after disease progression (required unblinding on patient level after documented disease progression)

Motivation (2)
RECORD-1: PFS and OS results

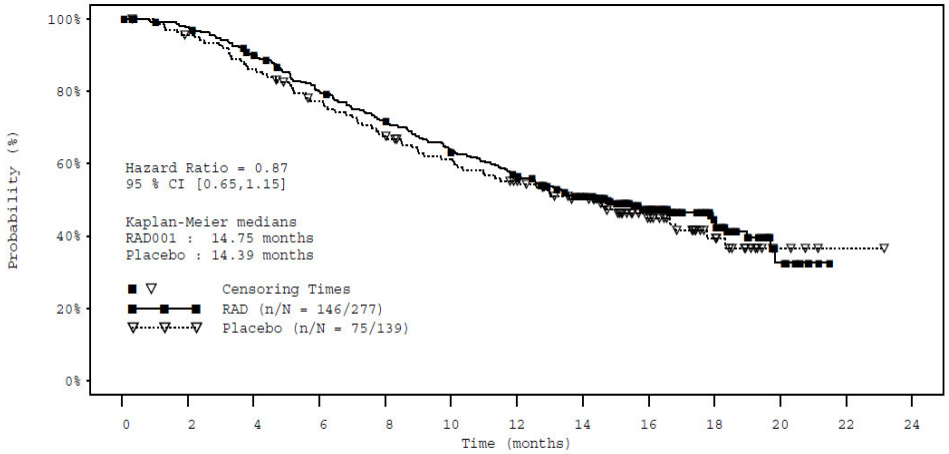
Progression-free survival (PFS)



No. of patients still at risk

Time (months)	0	2	4	6	8	10	12	14
RAD001	277	192	115	51	26	10	1	0
Placebo	139	47	15	6	2	0	0	0

Overall survival (OS)

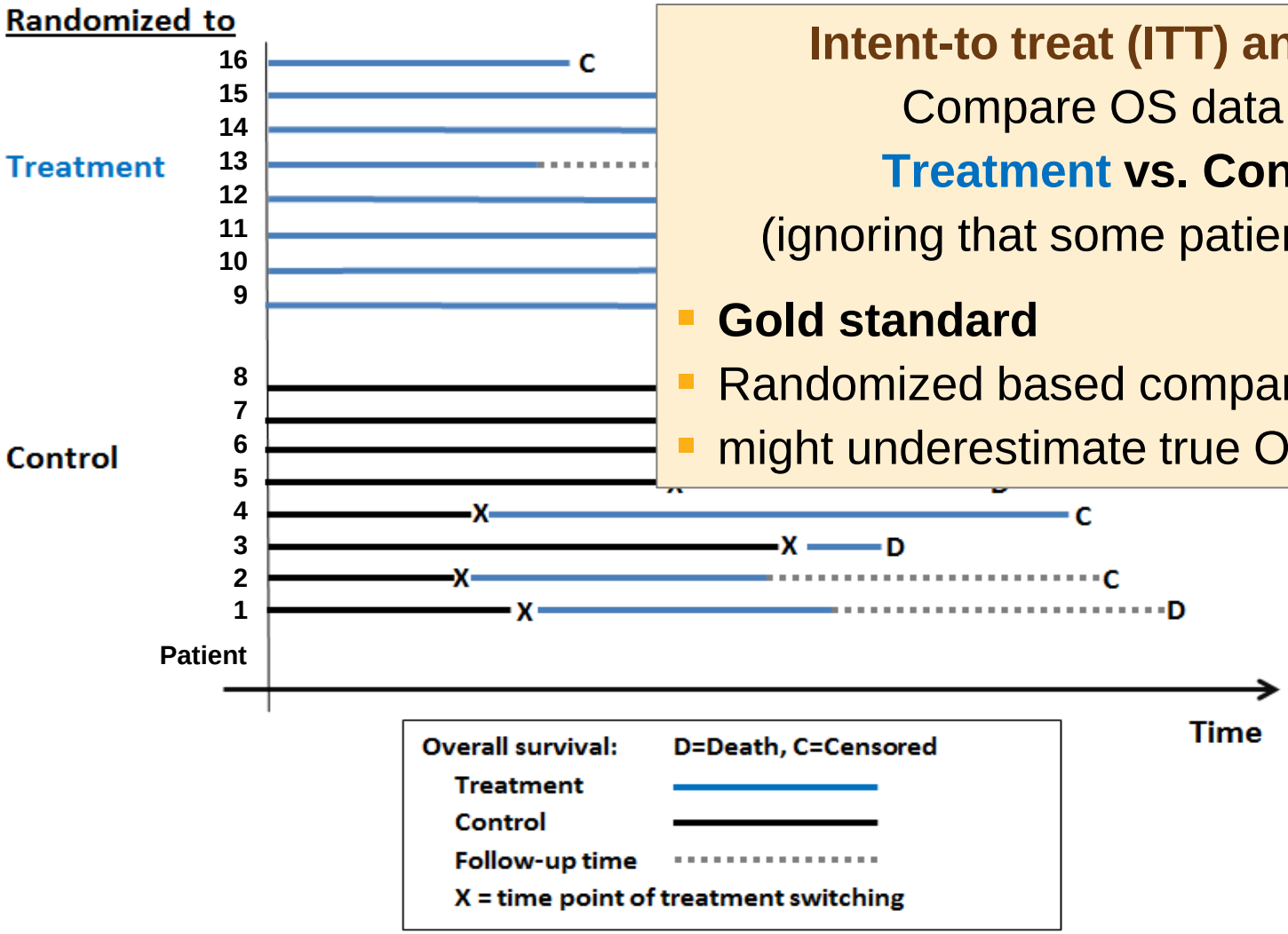


No. of patients still at risk

Time (months)	0	2	4	6	8	10	12	14	16	18	20	22	24
RAD	277	267	243	213	190	166	147	121	77	41	8	0	0
Placebo	139	131	118	103	89	79	68	57	33	17	4	1	0

- PFS comparison: HR=0.33 [0.25;0.43]; logrank test p-value < 0.0001 medians PFS 4.9 vs. 1.9 months
- 111/139 (80%) of Placebo patients switched to open-label everolimus (most at week 8 or 16) => very likely to confound intent-to-treat analysis of OS
- Overall survival: 221 deaths observed, p-value=0.162, HR=0.87 [0.65-1.15]

Illustration of data for overall survival (OS) in a study with treatment switching



Intent-to treat (ITT) analysis:

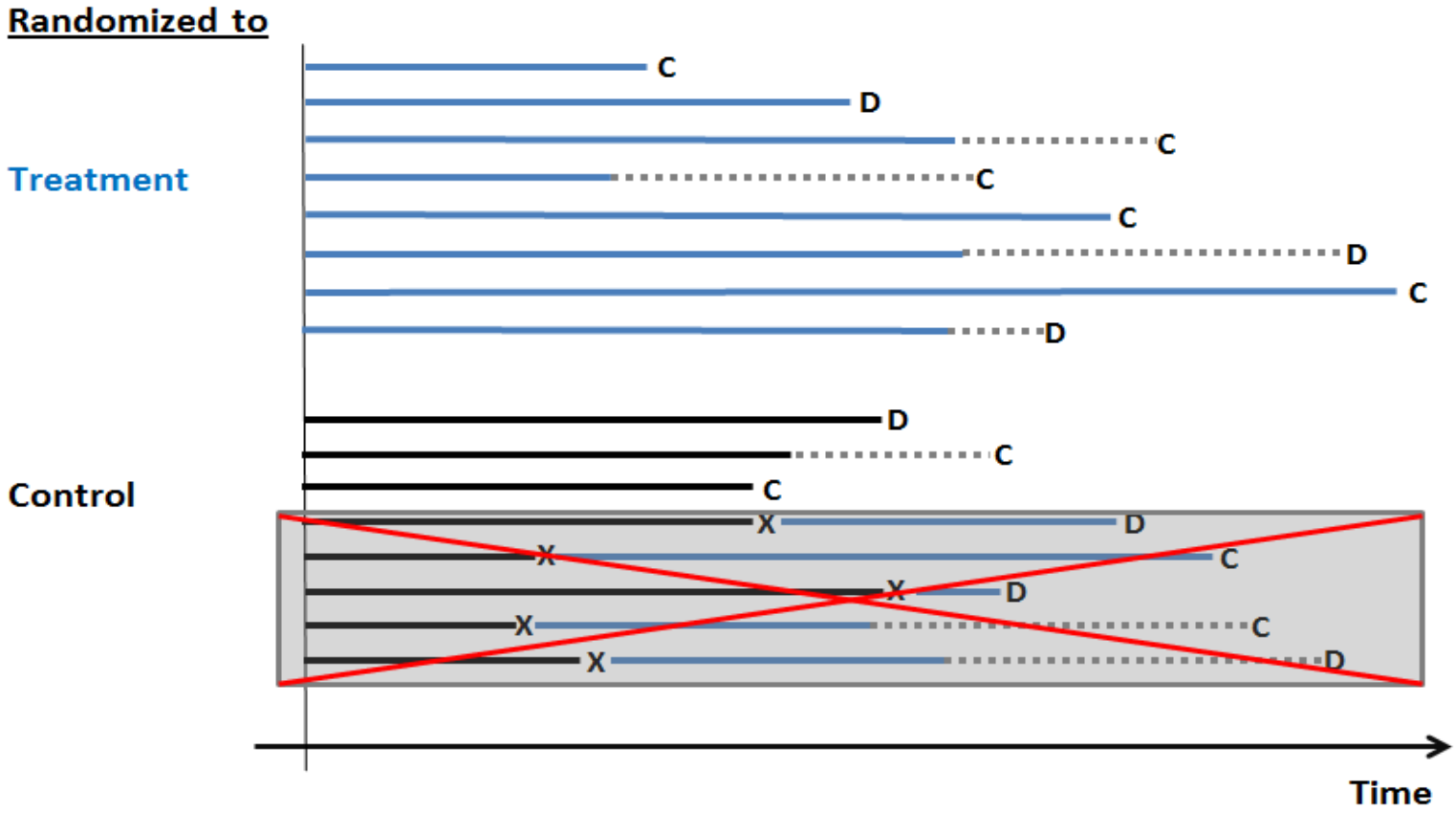
Compare OS data for

Treatment vs. Control

(ignoring that some patients switched)

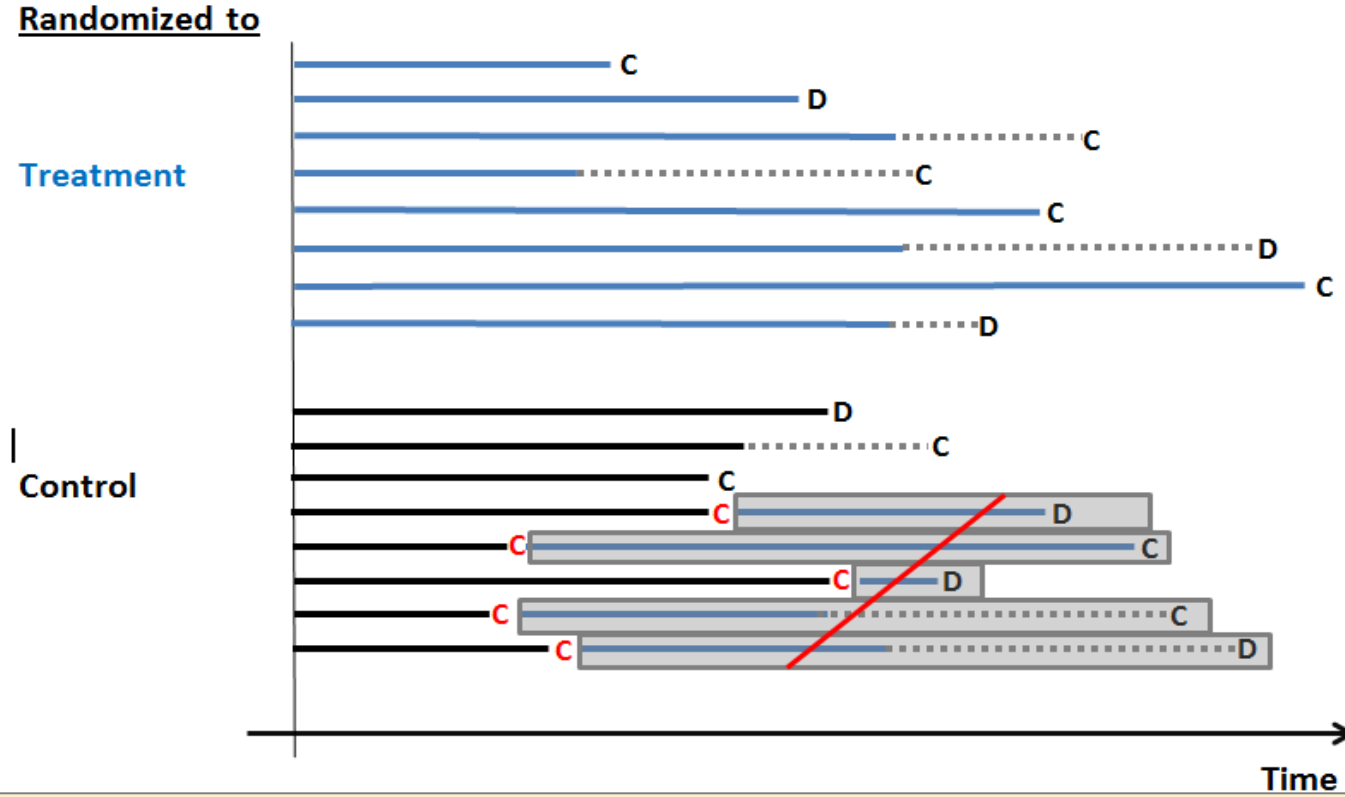
- **Gold standard**
- Randomized based comparison
- might underestimate true OS difference

Simple methods: Exclude treatment switchers



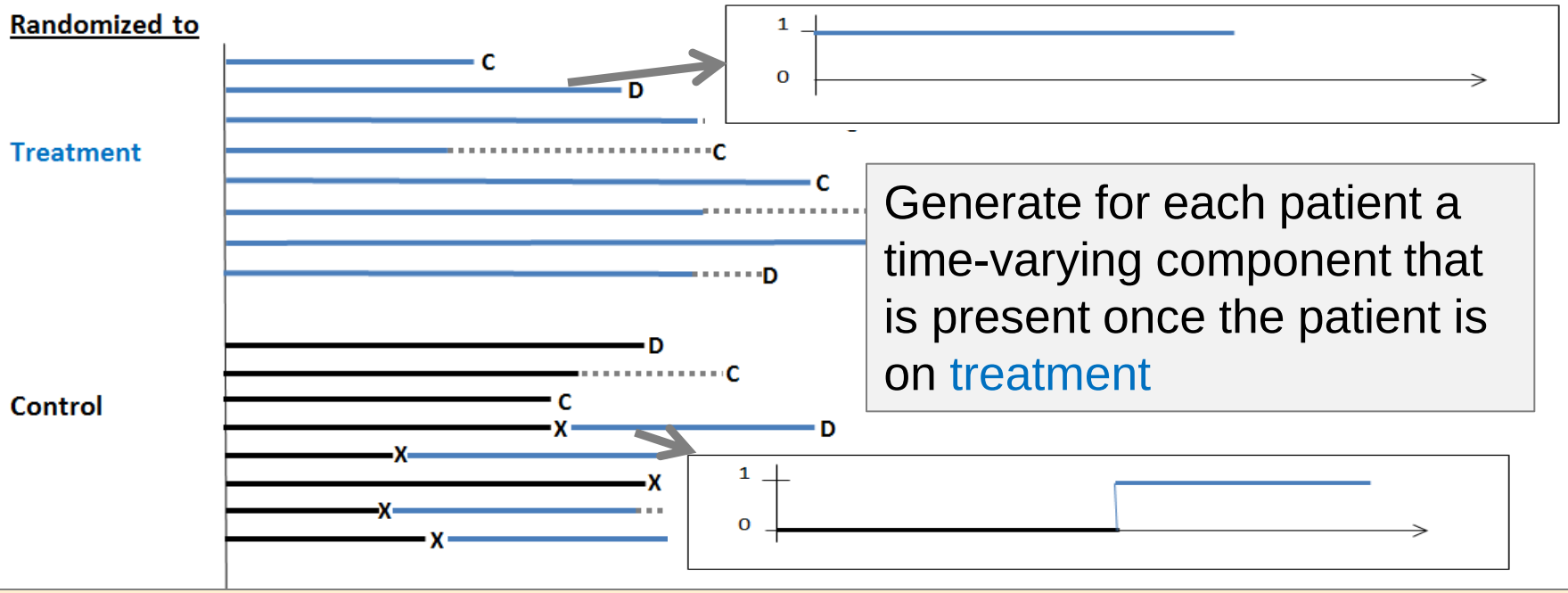
- Excludes patients from Control only
- no comparison of randomized groups
- Might produce heavily biased results => **not a good idea at all**

Simple methods:
Censor switchers at ‘time point of cross-over’



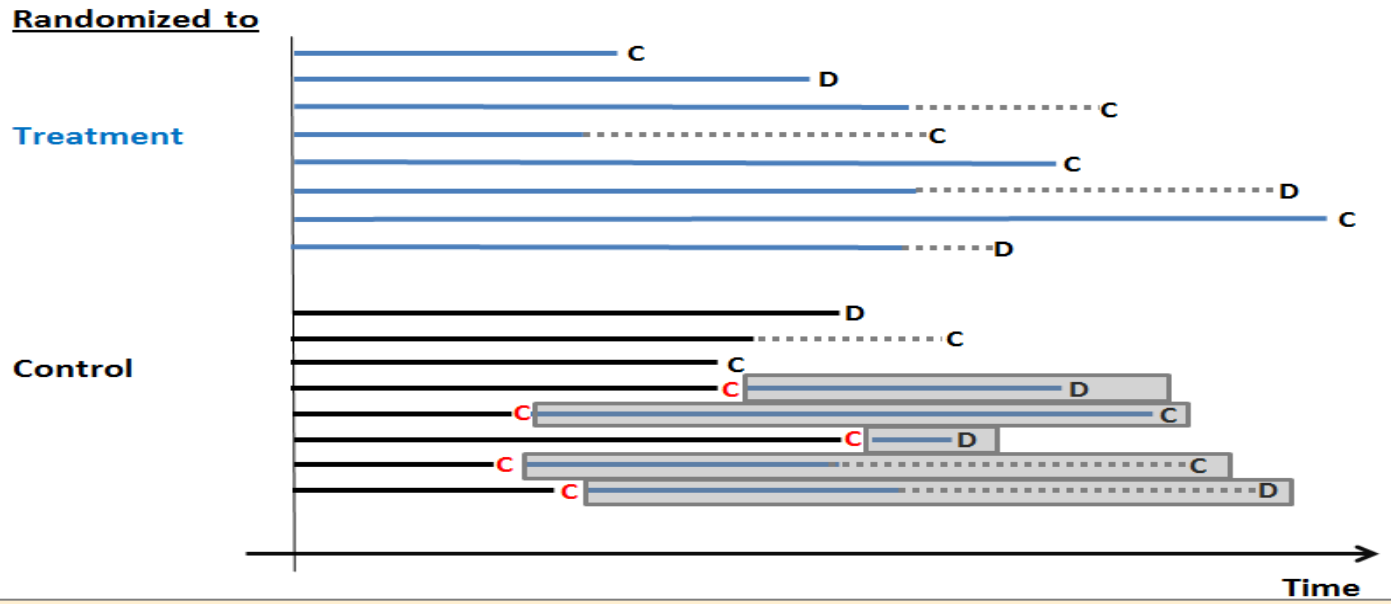
- Simple approach, compares randomized groups (RECORD-1: HR=0.76, 95%CI [0.46, 1.27])
- Reasonable if the OS prognosis of patients who switched treatment is equal to those who did not switch => likely to be violated (e.g. Patients with poor prognosis more likely to switch) => informative censoring, results biased

Simple methods:
Use a time-varying treatment variable



- Estimate the treatment effect by including the time-varying component in a regression model (e.g. Cox proportional hazards model)
- No longer a comparison between randomized **Treatment** **vs. Control** arm, more difficult interpretation
- 'no counfounders assumption', might be biased in case of other time-depenedent influence factors (e.g. OS prognosis might be worse after disease progression/treatment switch)

Complex methods: Inverse-probability-of-censoring weighting (IPCW)



- Switchers are censored at ‘time point of cross-over’, but patients are **weighted** according to their probability to switch treatment.
- IPCW method artifically
  - **increases** weights for patients with **low probability** of treatment switch
  - **decreases** weights for patients with **high probability** of treatment switch

IPCW – steps for data preparation and analysis

Data preparation

1. identify baseline covariates & time dependent confounders
2. create data panel (i.e. split follow-up period in time intervals with matching patient status and covariates)

Key assumption for IPCW: ***no unmeasured confounders***

i.e. all baseline covariates and all post-baseline time dependent confounders that predict both, treatment switch and outcome OS, are included => ***questionable that this is always fulfilled***

IPCW analysis (2 steps)

A. Determine IPCW weights (e.g. via logistic regression model)
B. Apply resulting weights in the analysis of Overall survival (e.g. weighted Cox regression model)

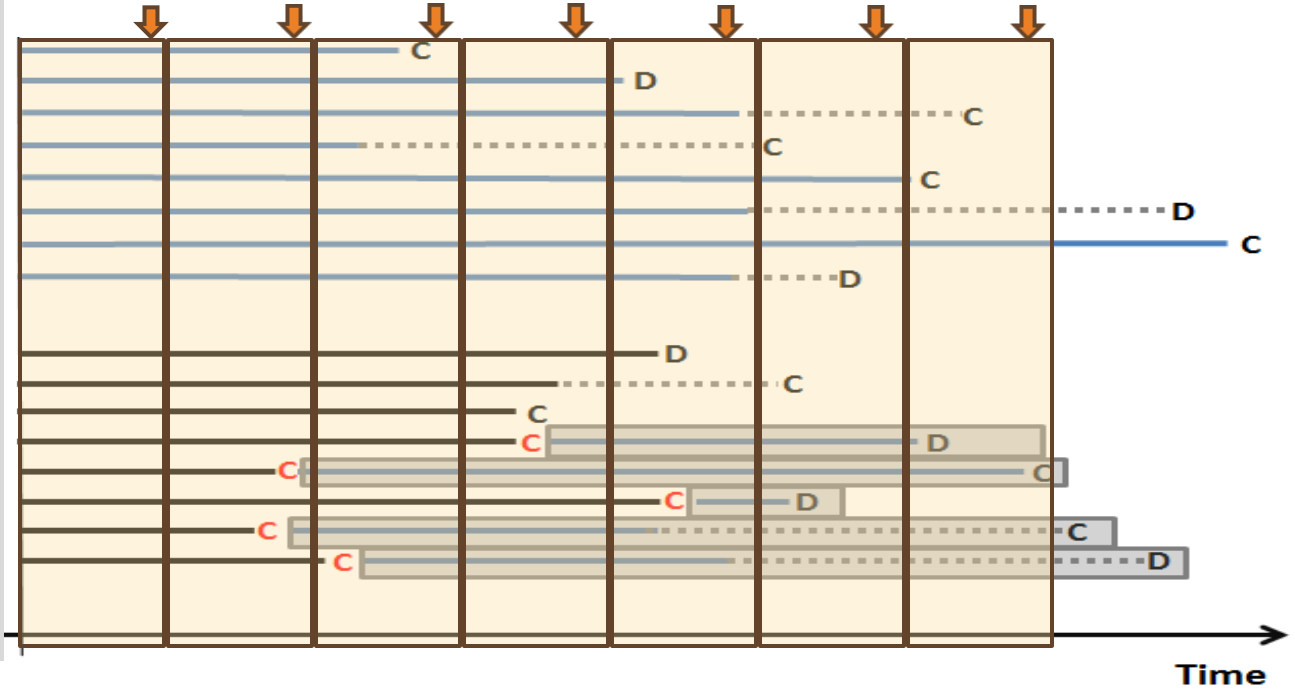
Practical question:

Which covariates to be included: All or 'relevant covariates' only ?

IPCW – application to the RECORD-1 study

Time-dependent confounders
Time period, **KPS**, assessment of **disease progression** (local)

Baseline covariates
-**randomized treatment**
-country
-age (years)
-sex
-race
-MSKCC prognostic score at baseline
-KPS at baseline
-prior sorafenib only
-prior sunitinib only
-prior treatment with both
-prior radiation treatment
-prior nephrectomy
-time since diagnosis
-liver involvement
-bone involvement



Step A: Determine IPCW weights

The final logistic regression model included

- baseline measures (age, sex, prior VEGF treatment, and baseline MSKCC score)
- time-updated covariates (KPS post-baseline and progression status)

IPCW – Step B: Weighted Cox regression models for OS in RECORD-1 (sensitivity analyses)

Table 1. Variables Included in All Cox Regressions Models Considered

Description	Model 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Age at baseline (y)	✓		✓	✓		✓								✓	✓
Country			✓		✓										
Sex	✓		✓	✓			✓							✓	✓
Race			✓	✓				✓						✓	✓
MSKCC prognostic score at baseline	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
KPS at baseline															✓
Prior treatment with sorafenib only	✓		✓	✓							✓	✓	✓		✓
Prior treatment with sunitinib only	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Prior radiation treatment			✓	✓					✓				✓		✓
Prior nephrectomy			✓	✓						✓			✓		
Time since diagnosis			✓	✓											✓
Liver involvement		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Bone involvement		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Randomized treatment	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Time period	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
HR	0.54	0.49	0.45	0.47	0.50	0.49	0.53	0.49	0.50	0.51	0.44	0.49	0.51	0.52	0.49
HR 95% CI	0.30, 1.01	0.26, 0.92	0.24, 0.84	0.27, 0.82	0.27, 0.94	0.26, 0.91	0.28, 1.00	0.26, 0.91	0.27, 0.93	0.27, 0.96	0.26, 0.76	0.26, 0.92	0.27, 0.96	0.28, 0.98	0.29, 0.83

CI=confidence interval; HR=hazard ratio; KPS=Karnofsky performance status; MSKCC=Mem

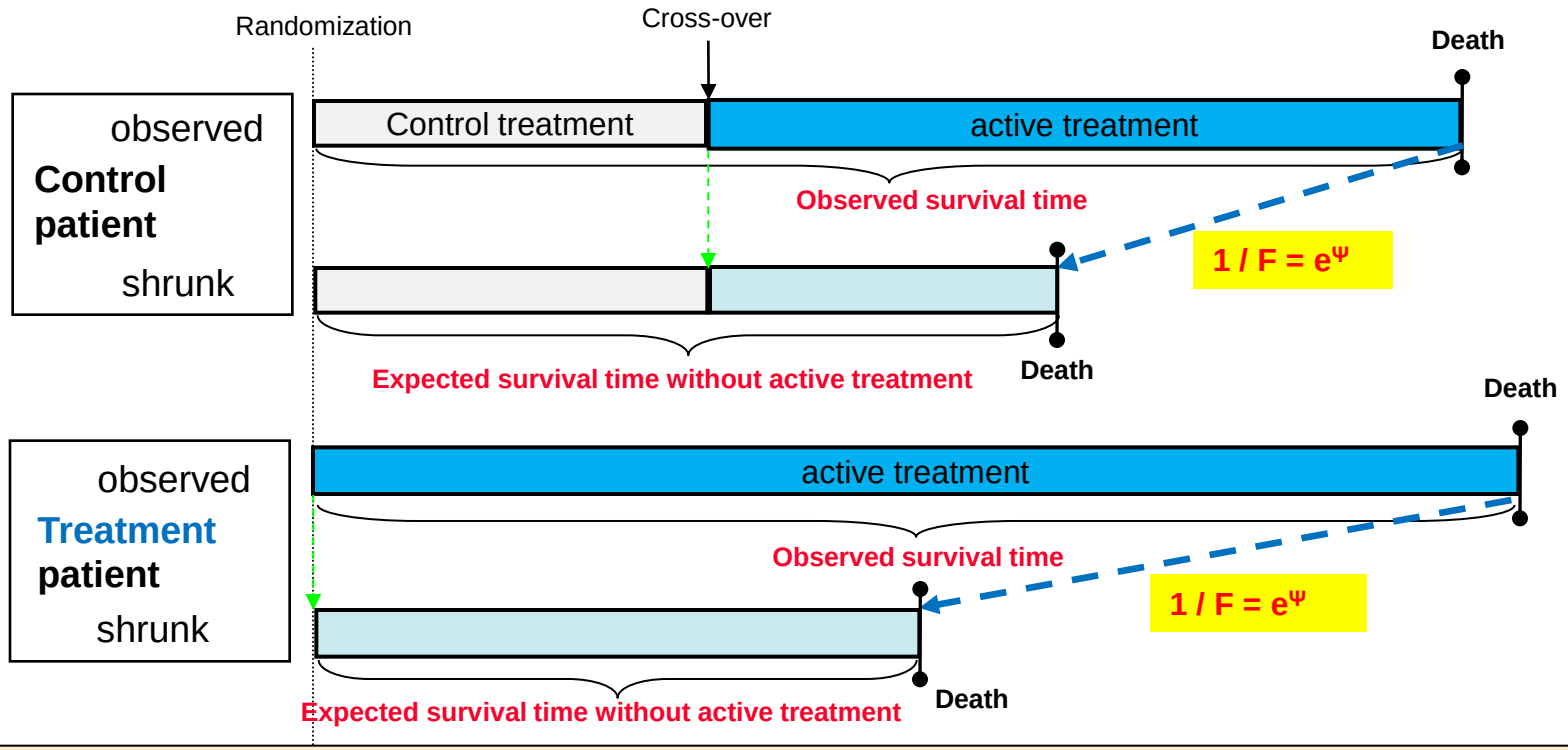
Model 4: best model fit (AIC)

Complex methods:
Rank Preserving Structural Failure Time Model (RPSFT)

Key principles

- Estimate the **survival time gained/lost by receiving active treatment** (i.e. either randomized or “cross-over” active treatment)
- Main assumption: **treatment is acting by multiplying survival time by a given factor once patient starts receiving active treatment** (**transparent but un-testable assumption**)
- Multiplicative factor interpreted as relative increase/decrease in survival if one took active treatment compared to taking control
- It works by reconstructing the survival duration of patients, as if they had never received active treatment

RPSFT – ‘Shrinking’ Survival Times

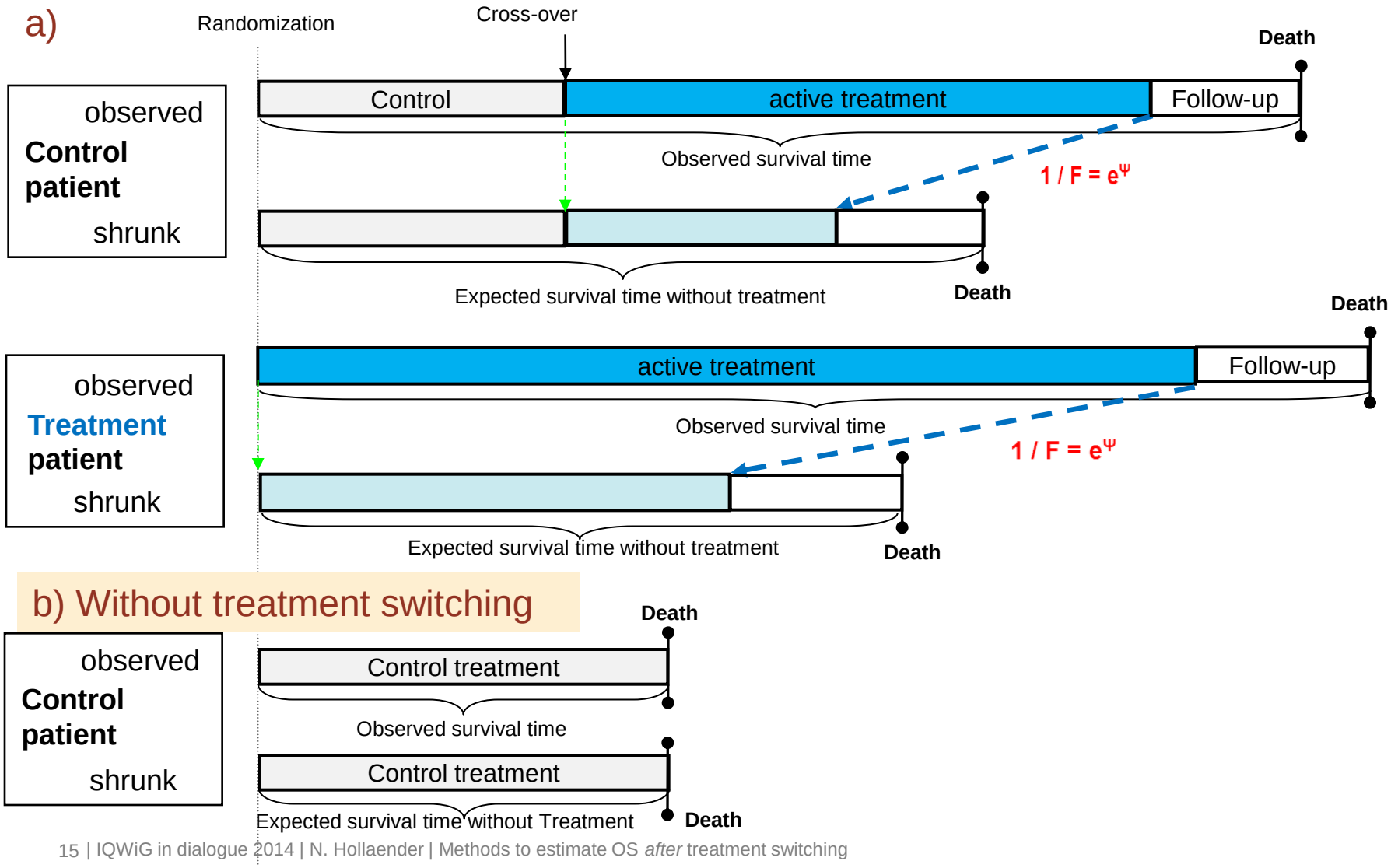


Grid search for factor F:

repeat ’shrinking’ of time on treatment for all patients by varying a factor F (via Ψ) until both survival curves (test and control) can no longer be distinguished, i.e. as if all patients only received placebo

RPSFT – ‘Shrinking’ of Survival Times:

a) with follow-up after treatment discontinuation
b) in control arm without treatment switching



RPSFT – application to RECORD-1 (based on logrank test)

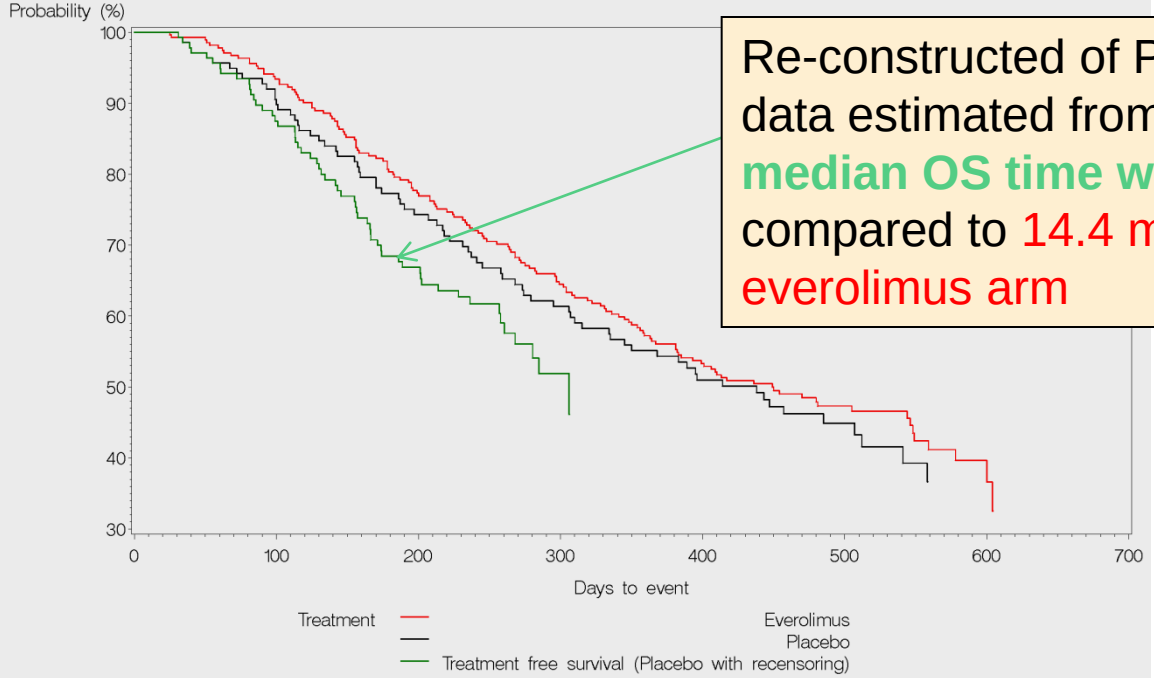
- Result $\Psi^* = -0.66$ obtained when selecting the most conservative point over a grid from -2.50 and 1 in steps of 0.01

- F = 1.93 with 95% CI (0.50; 8.50), i.e the estimated **relative survival time** for patients treated (always) with everolimus is **1.93 times longer** as compared to patients never treated with everolimus (i.e. placebo without crossover)

RPSFT

- provides a randomization based treatment effect estimator
- assumes that treatment effect is the same regardless of when the experimental treatment is initiated (might be extended to allow different effect before/after switching)
- extra censoring required to maintain the assumption of independent random censoring (=> reduces precision)

RPSFT – RECORD-1
Re-constructed KM curve for Placebo arm



Re-constructed of Placebo arm based on data estimated from RPSFT model, **median OS time would be 10.0 months** compared to 14.4 months in the observed everolimus arm

Convert RPSFT results to HR scale:

HR=0.60 with 95%CI (0.22; 1.65)

Attention: to estimate the CI correction is required to account for model selection

Summary and overview
(see also Watkins et al., Pharmaceutical Statistics, 2013)

Method	Description	Assumption(s)	Limitations/Comments
Censoring at crossover date (HR scale)	simple naïve approach, OS time censored at the time of switching/start of experimental treatment.	Censoring must be non-informative	Often informative censoring, therefore biased estimate of treatment effect
Time-varying Cox PH model (HR scale)	relatively simple, considers treatment as a time-varying covariate	Delayed treatment has the same effect on survival as treatment started upfront. No time-dependent confounding factors present	Results often biased in the presence of confounding factors, i.e. time dependent covariates that that affects OS and treatment switching
Inverse probability censoring weighting (IPCW) (HR scale)	provides unbiased estimate of treatment effect on OS given the all baseline and time-dependent covariates are correctly specified.	No unmeasured confounders (all factors influencing crossover and survival are included in the model).	• Assumption on absence of unmeasured confounders is untestable. • Method is not applicable if there are very few patients who did not switch and experienced an event. • Results might be sensitive to model building steps
Rank preserving structural failure time (RPSFT) model (time scale)	Model-based method that reconstructs artificial survival time in the absence of experimental treatment.	Treatment effect is the same regardless of when the experimental treatment is initiated, e.g. delayed start of experimental treatment has the same effect as starting upfront.	• The structural assumption is untestable. • Results are sensitive to the method used for determination of acceleration factor F. • Re-censoring applied to all censored patients irrespectively of switch.

Conclusions

- ITT analysis is gold standard – but completely ignoring heavy treatment switch not recommended (underestimates OS benefit)
- There is no best method to correct the OS comparison for treatment switching, all methods have pros and cons
- Complex methods are more appropriate

Points to consider:

- How many patients switched treatment?
- When did the treatment switch occurr?
- Are assumptions required for a specific methods reasonable?
- Describe details of applied methods
- How stable are the results? (=> sensitivty analyses)
- ... other

References (related to RECORD-1 study)

- Motzer RJ, Escudier B, Oudard S, Hutson TE, Porta C, Bracarda S, Grünwald V, Thompson JA, Figlin RA, Hollaender N, Urbanowitz G, Berg WJ, Kay A, Lebwohl D, Ravaud A; RECORD-1 Study Group. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. Lancet. 2008 Aug 9;372(9637):449-56. Epub 2008 Jul 22.
- Motzer RJ, Escudier B, Oudard S, Hutson TE, Porta C, Bracarda S, Grünwald V, Thompson JA, Figlin RA, Hollaender N, Kay A, Ravaud, A: Phase 3 Trial of Everolimus for Metastatic Renal Cell Carcinoma: Final Results and Analysis of Prognostic Factors, Cancer, 2010; 116 (Issue 18): 4256-4265.
- Korhonen P, Malangone E, Sherman S, Casciano R, Motzer RJ, Baladi J, Haas T, Zuber E, Hollaender N, and Lebwohl D: Overall Survival of Metastatic Renal Cell Carcinoma Patients Corrected for Crossover Using Inverse Probability of Censoring Weights and Rank-Preserving Structural Failure Time Models: Two Analyses From the RECORD-1 trial. 2010 ASCO Annual Meeting, June 4 – 8, 2010, Chicago, IL, Abstract #4595, Poster Presentation.
- Korhonen P, Zuber E, Branson M , Hollaender N , Yateman N , Katiskalahti T, Lebwohl D, and Haas T: Correcting Overall Survival for the Impact of Crossover Via a Rank-Preserving Structural Failure Time (RPSFT) Model in the RECORD-1 Trial of Everolimus in Metastatic Renal-Cell Carcinoma. Journal of Biopharmaceutical Statistics, 2012; 22(6): 1258-1271

References (a few papers on methods

- **Watkins C, Huang X, Latimer N, Tang Y and Wright EJ. Adjusting overall survival for treatment switches: commonly used methods and practical application. Pharmaceutical Statistics, 2013; 12(6): 348-57.**
- Morden JP, Lambert PC, Latimer N, Abrams KR and Wailoo AW. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. BMC Medical Research Methodology 2011, 11:4
- Cole SR, Hernan MA. Constructing Inverse Probability Weights for Marginal Structural Models. American Journal of Epidemiology. 2008; Vol. 168, No. 6.
- Hernan MA, Cole SR, Margolick J, Cohen M and Robins JM. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. Pharmacoepidemiology and drug safety 2005; 14: 477–491
- White IR, Babiker AG, Walker S, Darbyshire JH. Randomization-based methods for correcting for treatment changes: examples from the Concorde trial. Stat Med. 1999 Oct 15;18(19):2617-34.
- ...

BACK-UP

SLIDES

Structural Model used for RPSFT

U_i potential treatment free survival time (would have been observed if no RAD001 had been received)

$$U_i = \int_0^{T_i} e^{\psi A_i(s)} ds$$

where $A_i(t)$ = 1 if patient received RAD001 at time t, and 0 otherwise.
Here: once a patient discontinues treatment he/she never starts treatment again, the model can be simplified as

$$U_i(\psi) = \int_0^{A_i} e^{\psi \cdot 1} ds + \int_0^{T_i - A_i} e^{\psi \cdot 0} ds = A_i e^{\psi} + (T_i - A_i)$$

Structural Model used for RPSFT (cont'ed)

Survival time then given as

$$T_i = U_i - A_i\left(e^{\psi} - 1\right)$$

Time gained (lost) while on active treatment

If always on active treatment RAD001 ($A_i = T_i$):

$$U_i = e^{\psi} T_i \qquad T_i = e^{-\psi} U_i$$

If never on active treatment RAD001 ($A_i = 0$):

$$U_i = T_i$$

RPFST postulates that each day spent on RAD001 prolongs (reduces) the survival time by a multiplicative fator $e^{-\psi}$

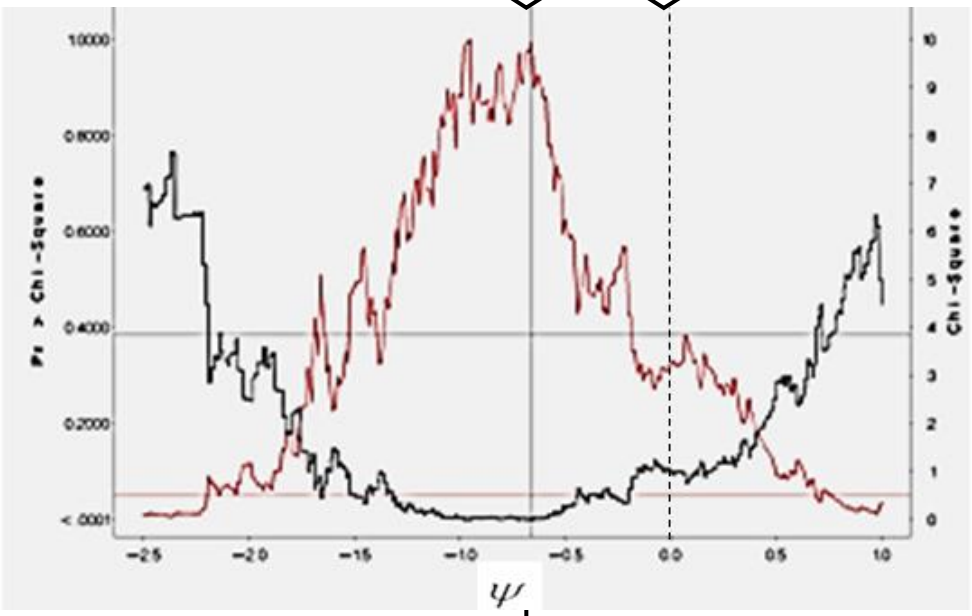
Task: estimation of ψ and therefore of $e^{-\psi}$

(e.g. negative values of ψ indicate longer survival when treated)

RPSFT - Estimation procedure (RECORD 1 study)

Selected estimate $\Psi^* = -0.66$
95% CI (-2.14; 0.69)

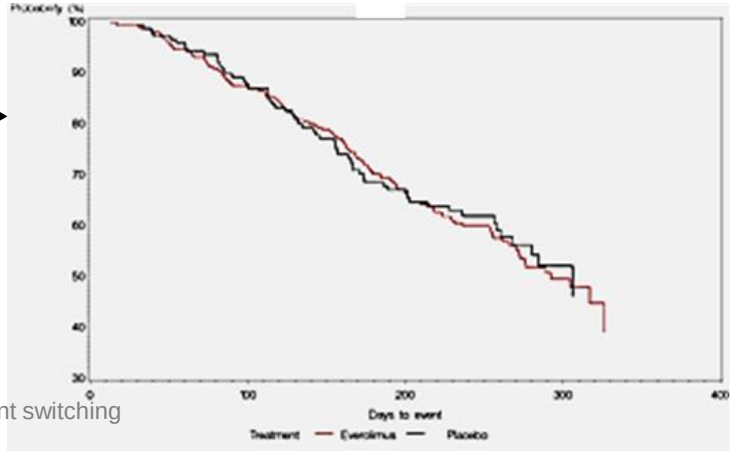
$\Psi = 0$ would correspond to ITT analysis



The point estimate of ψ is found as the point where the **log rank test statistic (black line)** is at its **minimum** or equivalently the **p-value (red line)** is at its **maximum**.

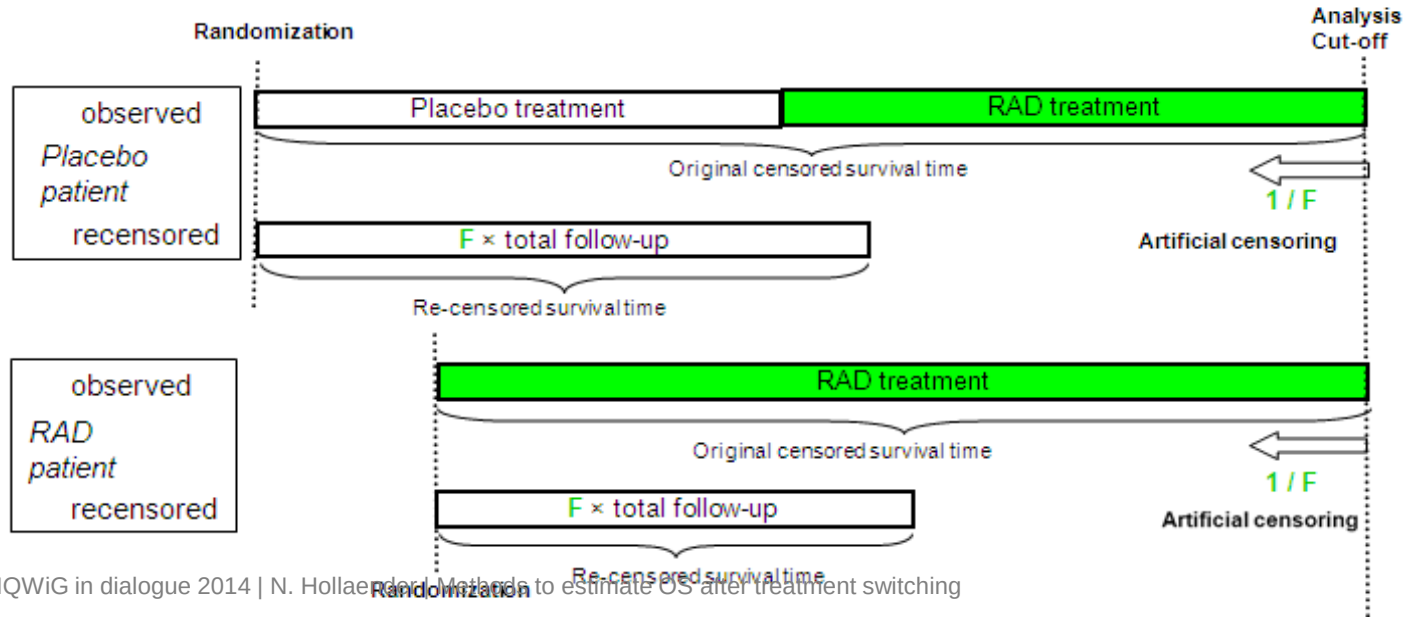
95% CI bounds are found as points where the log rank test statistic hits 3.84 (upper horizontal line) or equivalently where p-value hits 0.05 (lower horizontal line).

Resulting survival curves for $\Psi^* = -0.66$ do not indicate any difference between treatment arms



RPSFT – Artificial censoring algorithm

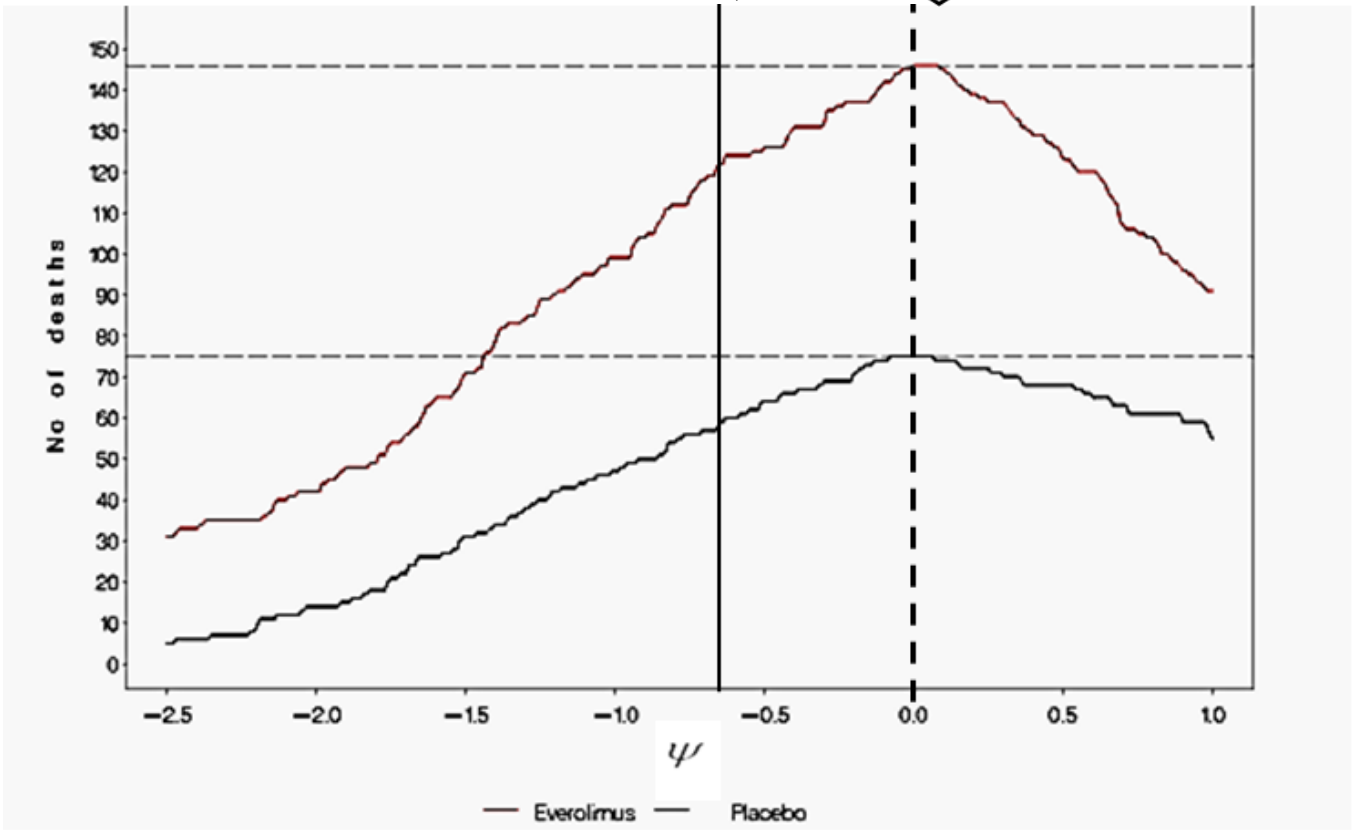
- An additional algorithm (‘artificial-censoring’) allows to **maintain the assumption of independent random censoring required for unbiased estimation**
- The artificial censoring algorithm works by shrinking the total follow-up time (time between randomization to analysis cut-off date) for all patients regardless of randomization group or treatment received
- Therefore every patient censored in the ITT analysis remains censored with duration equal or shorter to the original one; in addition, patients with an event in the original analysis may become censored via the artificial-censoring algorithm



Impact of artificial censoring on number of events (deaths) used in RPSFT

For $\Psi = 0$ (ITT analysis):
146 deaths in RAD001 arm
75 deaths in palcebo arm

$\Psi^* = -0.66$



Attention: Extra censoring reduces precision !

Hazard ratio of the ‘corrected treatment effect’ on OS

- The hazard ratio
  HR(observed RAD001 arm vs. Re-constructed placebo arm)
  can be estimated in a Cox proportional hazards model

- **Attention**: Do not use naive estimates of standard deviation sd, 95% CIs and p-value obtained in PHREG

- For estimation of ‘corrected’ standard deviation sd*: increase naive estimated sd standard deviation by inflation factor to reflect the p-value obtained in the ITT analysis, i.e. use sd* = sd · inflation factor

- In RECORD-1 we obtained
  HR=0.60 with 95%CI (0.22; 1.65)

Final remarks (further topics)

- Estimate $\Psi^* = -0.66$ was based on selecting the most conservative point over a grid from -2.50 and 1 in steps of 0.01, finer grid might lead to other point estimates but there was hardly any impact on CI
- Results presented in this talk based on logrank-test, sensitivity analysis based on other test statistics (Fleming-Harrington G^p-family) provided in the forthcoming paper
- Model might be extended to allow the treatment effect to be different before and after cross-over from Placebo (see discussion of forthcoming paper)

RPSFT (Sensitivity analysis for RECORD-1, see Korhonnen et al, 2012)

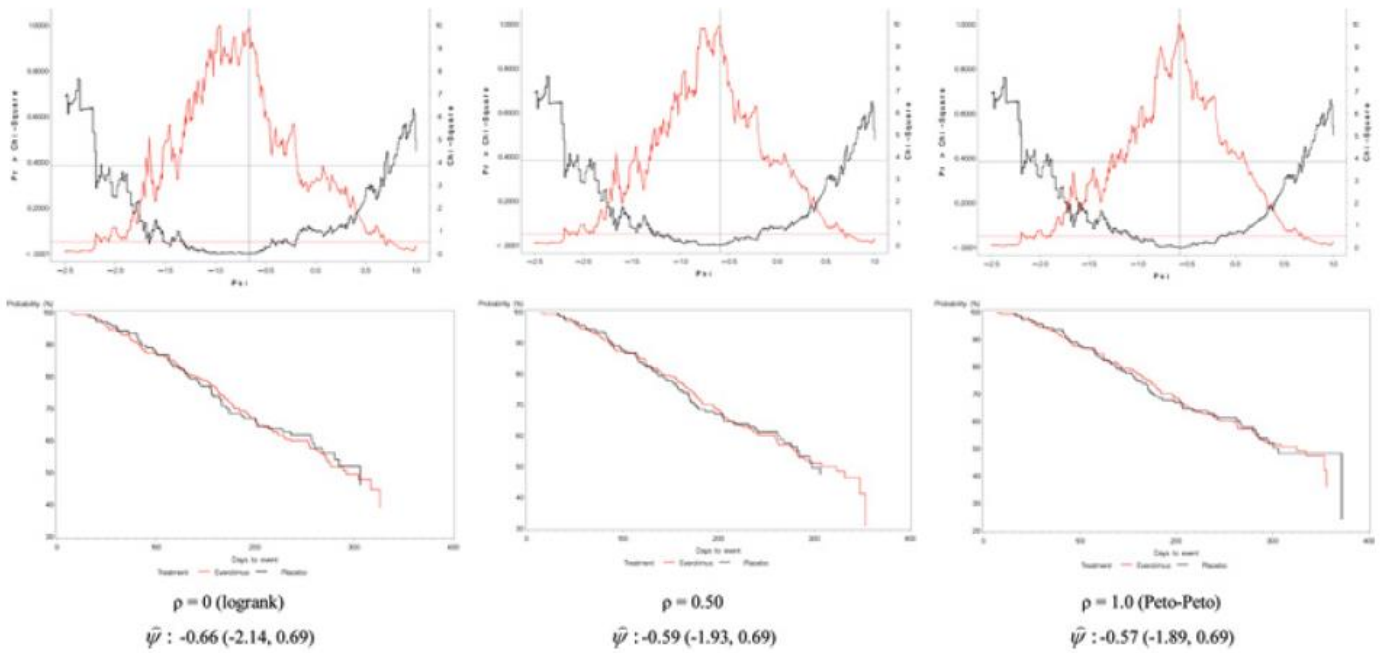


Figure 2 Estimation of structural parameter ψ for 3 different Fleming–Harrington G^{ρ} statistics ($\rho = 0, 0.5$, and 1). Upper panel shows distribution of test statistic (black line) and p-value (red line); lower panel presents reconstructed treatment-free Kaplan–Meier survival curves for each treatment arm using $\widehat{\psi}$. (Color figure available online.)

RPSFT (Sensitivity analysis for RECORD-1, see Korhonnen et al, 2012)

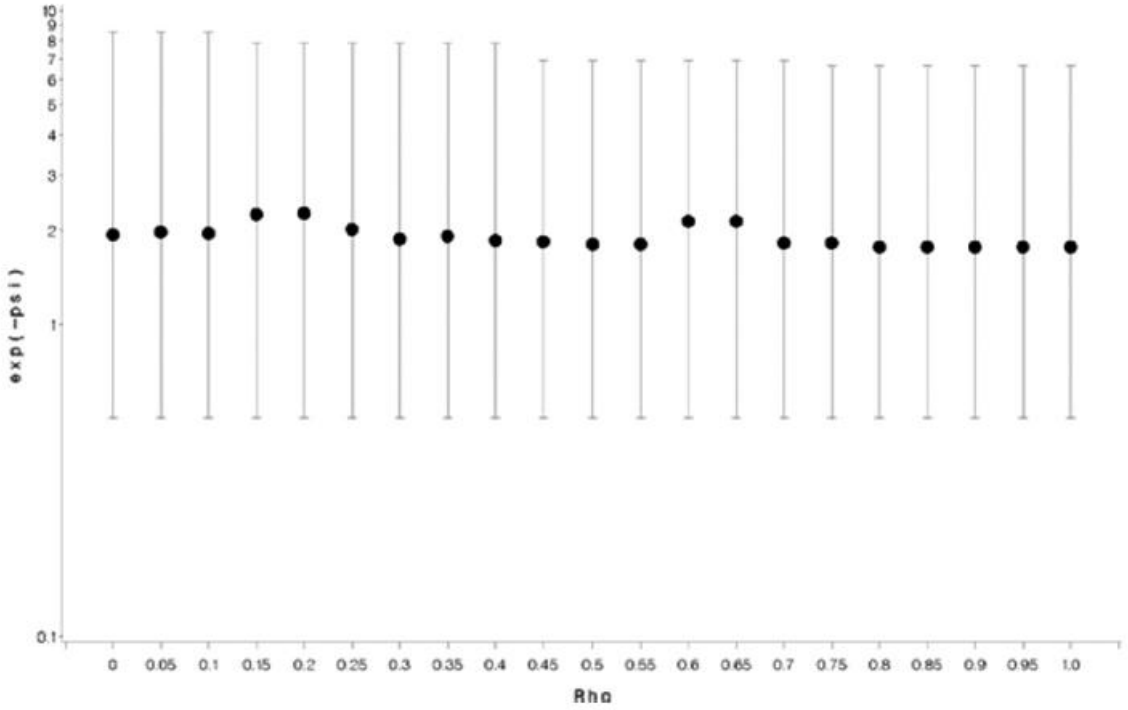


Figure 3 Estimated relative survival benefit, $\exp(-\widehat{\psi})$, with 95% confidence intervals for various values of ρ (rho) between 0 and 1.