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
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)

IQWiG in dialogue 2014 | N. Hollaender | Methods to estimate OS after treatment switching
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

Overall survival: D=Death, C=Censored
- Treatment
- Control
- Follow-up time
- X = time point of treatment switching
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 confounders assumption’, might be biased in case of other time-dependent 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 artificially
  - **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)

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)

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*

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)
### Table 1. Variables Included in All Cox Regressions Models Considered

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

CI = confidence interval; HR = hazard ratio; KPS = Karnofsky performance status; MSKCC = Memorial Sloan Kettering Cancer Center.

**Model 4: best model fit (AIC)**
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
Grid search for factor F:
repeat ’shrinking’ of time on treatment for all patients by varying a factor F (via \( \Psi \)) until both survival curves (test and control) can no longer be distinguished, i.e. as if all patients only received placebo

\[ 1 / F = e^\Psi \]
RPSFT – ‘Shrinking’ of Survival Times:

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

**Diagram:**

- **Randomization**
- **Cross-over**
- **Control**
- **Active treatment**
- **Follow-up**
- **Observed survival time**
- **Expected survival time without treatment**
- **Death**

### a) With follow-up after treatment discontinuation

- **Control patient**
  - Observed
  - Shrunk

- **Active treatment**
  - Observed survival time
  - Expected survival time without treatment

### b) Without treatment switching

- **Control patient**
  - Observed
  - Shrunk

- **Active treatment**
  - Observed survival time
  - Expected survival time without treatment
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)
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
<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Assumption(s)</th>
<th>Limitations/Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Censoring at crossover date (HR scale)</td>
<td>simple naive approach, OS time censored at the time of switching/start of experimental treatment.</td>
<td>Censoring must be non-informative</td>
<td>Often informative censoring, therefore biased estimate of treatment effect</td>
</tr>
<tr>
<td>Time-varying Cox PH model (HR scale)</td>
<td>relatively simple, considers treatment as a time-varying covariate</td>
<td>Delayed treatment has the same effect on survival as treatment started upfront. No time-dependent confounding factors present</td>
<td>Results often biased in the presence of confounding factors, i.e. time dependent covariates that affects OS and treatment switching</td>
</tr>
<tr>
<td>Inverse probability censoring weighting (IPCW) (HR scale)</td>
<td>provides unbiased estimate of treatment effect on OS given the all baseline and time-dependent covariates are correctly specified.</td>
<td>No unmeasured confounders (all factors influencing crossover and survival are included in the model).</td>
<td>• 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</td>
</tr>
<tr>
<td>Rank preserving structural failure time (RPSFT) model (time scale)</td>
<td>Model-based method that reconstructs artificial survival time in the absence of experimental treatment.</td>
<td>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.</td>
<td>• The structural assumption is untestable. • Results are sensitive to the method used for determination of acceleration factor F. • Re-censoring applied to allensored patients irrespectively of switch.</td>
</tr>
</tbody>
</table>
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 occur?
-Are assumptions required for a specific methods reasonable?
-Describe details of applied methods
-How stable are the results? (=> sensitivity analyses)
-... other


References (a few papers on methods)

- ...
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

\[ T_i = U_i - A_i(e^\psi - 1) \]

Time gained (lost) while on active treatment

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

\[ U_i = e^\psi T_i \quad T_i = e^{-\psi} U_i \]

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 factor \(e^{-\psi}\)

Task: estimation of \(\psi\) and therefore of \(e^{-\psi}\)

(e.g. negative values of \(\psi\) indicate longer survival when treated)
RPSFT - Estimation procedure (RECORD 1 study)

The point estimate of $\psi^*$ 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).

Selected estimate
95% CI (-2.14; 0.69)

$\psi^* = -0.66$

$\psi = 0$ would correspond to ITT analysis

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

$\Psi^* = -0.66$

Attention: Extra censoring reduces precision!
Hazard ratio of the ‘corrected treatment effect’ on OS

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

- **Attention**: Do not use naive estimates of standard deviation \( \text{sd} \), 95% CIs and p-value obtained in PHREG

- For estimation of ‘corrected’ standard deviation \( \text{sd}^* \):
  increase naive estimated \( \text{sd} \) standard deviation by inflation factor to reflect the p-value obtained in the ITT analysis, i.e. use \( \text{sd}^* = \text{sd} \cdot \text{inflation factor} \)

- In RECORD-1 we obtained
  \( \text{HR}=0.60 \) 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 Korhononnen et al, 2012)

Figure 2 Estimation of structural parameter $\psi$ for 3 different Fleming–Harrington $G^\rho$ statistics ($\rho = 0, 0.5, \text{and } 1$). 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 $\psi$. (Color figure available online.)
**Figure 3** Estimated relative survival benefit, $\exp(-\hat{\psi})$, with 95% confidence intervals for various values of $\rho$ (rho) between 0 and 1.