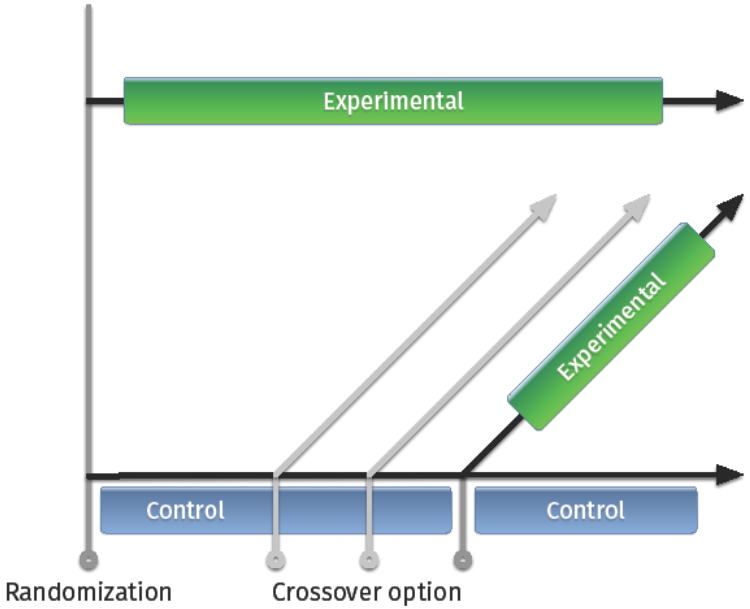
Getting closer to the truth?

Adjusting long-term efficacy estimates when patients cross over to experimental treatment

Dr. Sabine Lauer

PhUSE 2015

Crossover – the problem



- Common in Oncology trials where a strong efficacy signal early leads to an option for control arm patients to switch to the new therapy before later analyses
- Confounds intention-to-treat (ITT) analysis when patients are analyzed according to their randomized treatment
- In an ideal world, crossover should best not be allowed at all – but sometimes it needs to be introduced (e.g. ethical reasons)

Can we adjust for crossover in our analyses to get a better understanding of the magnitude of benefit?

Important note on data presented

All following considerations stem from the work on a "real" oncology trial but data had to be manipulated with random error for blinding purposes

The trial

- Main design features:
  - 1:1 randomization to experimental or control arm
  - Primary endpoint: disease-free survival (DFS)*
  - Secondary endpoints: overall survival (OS), safety
- Crossover option introduced after early positive interim analysis showing superiority of experimental vs. control for DFS
- High percentage of crossover (>30%)
- Patients in control arm with prior disease recurrence were NOT allowed to cross over

*DFS = Time from randomization to recurrence, second primary malignancy or death, whichever is first

Pre-planned analyses (DFS and OS)

- Primary analysis (requested as standard by authorities and payers):
  - ITT method
  - Conservative method and most important for drug approval
- Sensitivity analysis
  - Censor patients in the control arm at the time of crossover
  - Known to be biased as decision to crossover is usually not independent of prognostics
  - Bias may be in either direction(pro and contra experimental treatment)

Focus in the following will be on OS only as considerations for DFS are generally very similar

Results of pre-planned ITT and sensitivity analyses: Experimental vs. control

	Overall survival	
Method	HR	95% CI
ITT	0.80	(0.68, 0.94)
Censored	0.57	(0.48, 0.67)

Anti-conservative bias for the censored-method with bias in the set-at-risk in control arm to patients with poorer prognosis than the overall population (no crossover in case of prior recurrence)

Which other methods are currently known? (1)

- Naïve methods (not recommended due to high probability of bias)
  - Exclude switchers
    - Creates non-randomized subset with reduced sample size
    - Assumes that there are no confounders (variables influencing both survival and switching decision)
  - Time-varying covariate for treatment or switch
    - assumes again no “unmeasured” confounders (variables influencing both survival and the time-varying treatment variable)

Which other methods are currently known? (2)

- Complex methods
  - Inverse probability of censoring weighting (IPCW)
  - Rank-preserving structural failure time method (RPSFT)
  - Other less common methods, details not given here*

* For details, see the excellent overview article: Adjusting overall survival for treatment switches: commonly used methods and practical application, Watkins et al., Pharmaceutical Statistics 2013

IPCW (Inverse probability of censoring weighting)

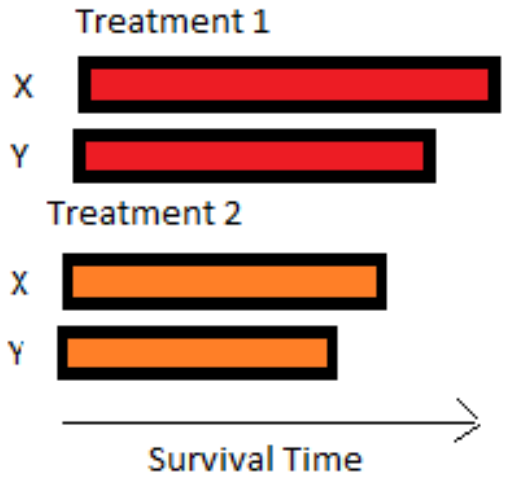
- Uses time-dependent Cox proportional hazard models with time-varying weights, giving larger weight for non-switch patients with similar characteristics to switchers
- Assumes no unmeasured confounders: all factors that influence both switch and survival are included in the weight calculation.
- IPWC not chosen for trial due to no prognostic variables assessed during the trial

Rank Preserving Structural Failure Time (RPSFT) Models

- Use concepts of latent survival time and counterfactual survival time:
  - A patient who switches treatment has a counterfactual event time – the time-to-event if no experimental treatment had been received
- Estimate the counterfactual event-time based on accelerated failure time models

RPSFT Models: Main assumptions

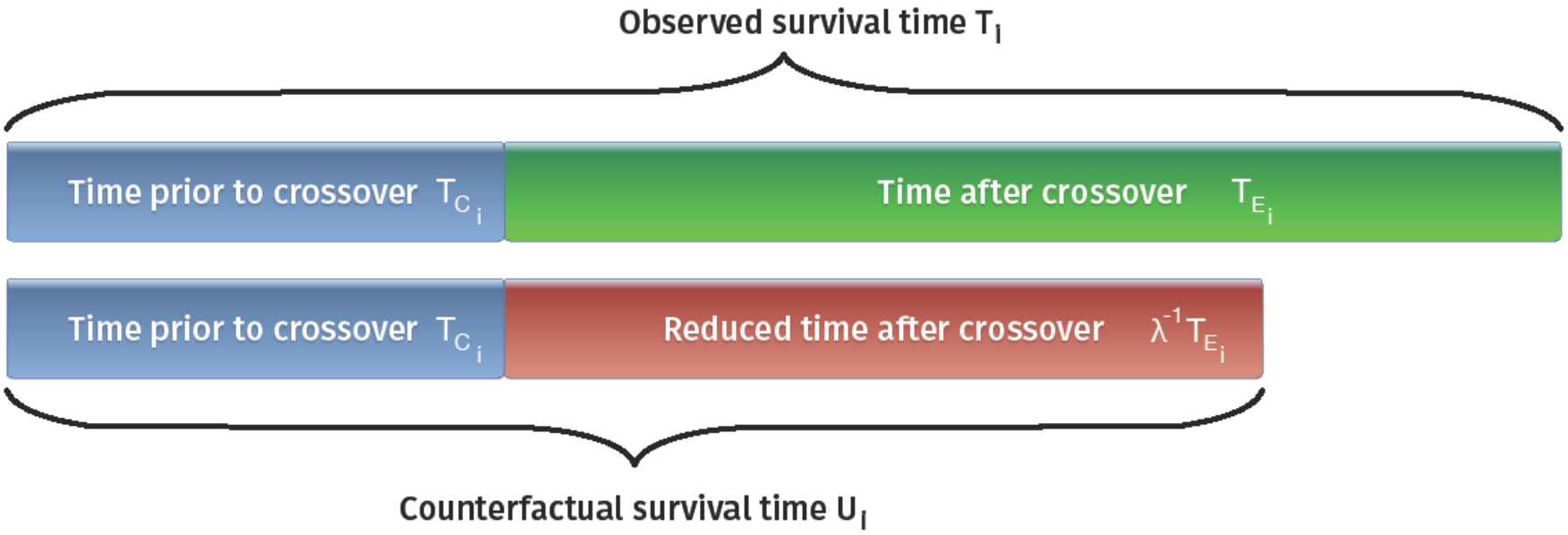
- Latent survival time is the same in both arms
- Rank-preserving: if patient X survives longer than patient Y with both receiving the same treatment, X would survive longer if both X and Y had received another (but still same) treatment



- Constant treatment effect i.e. same for first and second line treatment
- Mathematical model relating treatment exposure to survival

Chosen for trial (assumption of constant treatment effect presumably not heavily violated due to early introduction of crossover)

RPSFT – Calculating the counterfactual event time with a simple model



Assumptions (for all patients!):

$$T_i = T_{Ci} + T_{Ei}$$

$$U_i = T_{Ci} + \lambda^{-1} T_{Ei}$$

(for a suitable $\lambda > 0$, $\lambda > 1$ if treatment is beneficial)

How to estimate λ- test based grid search

U_i, the latent survival time, is a baseline characteristic and assumed to have the same distribution in each treatment arm

1. For all possible values of λ derive using the model $U_i(\lambda)$
2. Test if these $U_i(\lambda)$ are different between arms using any rank based test of the $U_i(\lambda)$ survival times (preferably using the same model as for the ITT analysis)
3. The value of λ that leads to no difference between the U_i in each arm is taken as the estimate of λ.

Note: there are also other approaches for estimation of λ (parametric approach suggested by Branson/Whitehead/hybrid approach suggested by Robins)

Handling Censored Data

- **Problem:**
  - **Survival time cannot be reduced in a similar fashion for censored patients without event**
  - **This would introduce bias due informative censoring: censoring dependent on time spent on experimental treatment (and thus treatment arm)**
- **Solution:**
  - **Re-censor all patients at the longest time they could have been observed for latent survival under each possible treatment history (using administrative cut-off date)**
  - **Note: Re-censoring means that we loose information**

Computational approach

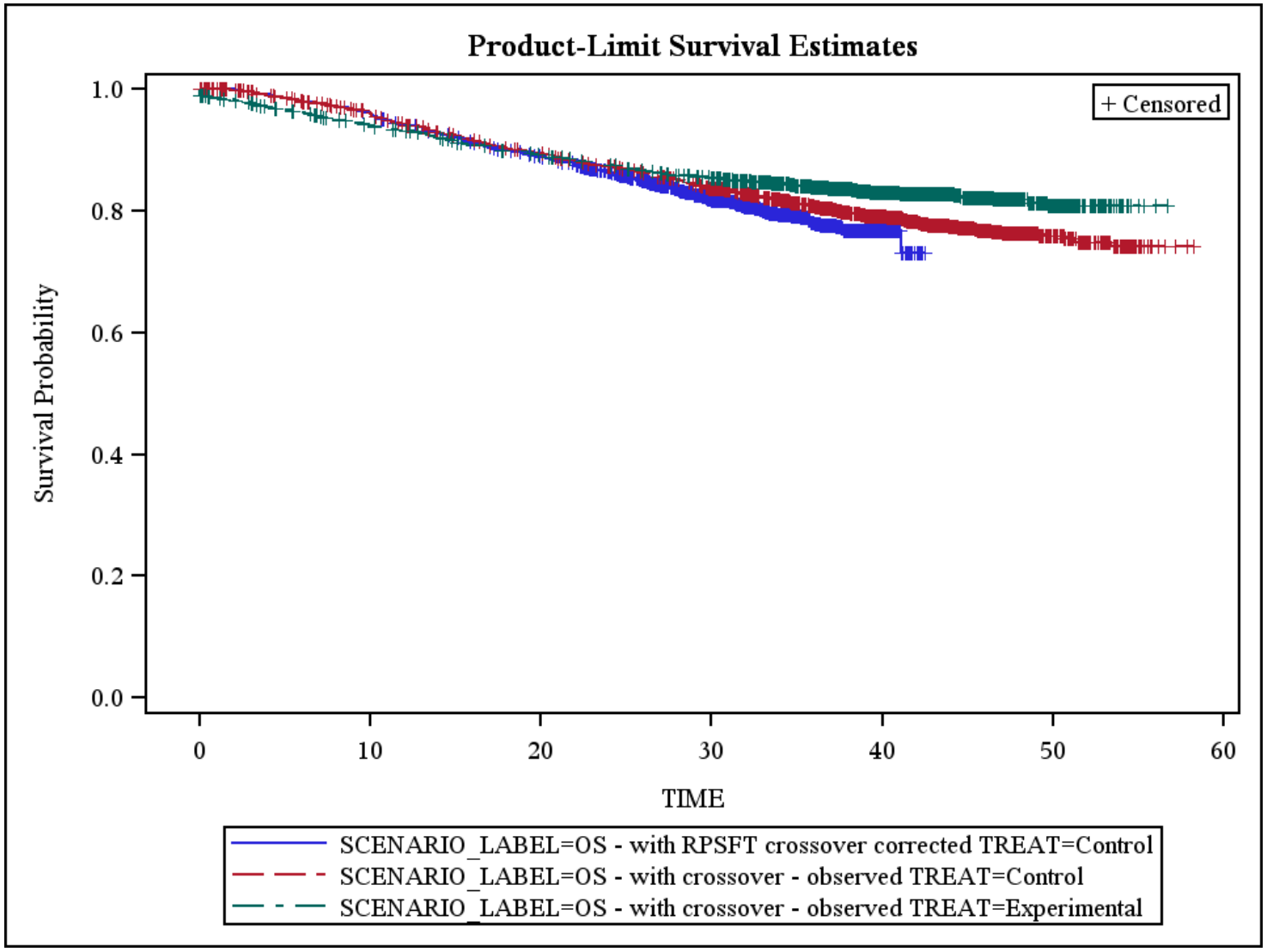
- Given the acceleration factor (λ) and time to administrative cut-off (CUT_i)
  1. In the control arm:
     1. Compute corrected survival times U_i for patients who crossed over (as previously described)
     2. For all patients, censor at $\lambda^{-1} CUT_i$ (in case of no earlier event/censoring)
  2. Compare these control arm data to the active arm T_i using usual standard techniques (log-rank test, Cox model, etc)
  3. Take point estimates in the usual way
  4. Derive estimates of uncertainty (e.g. SE, CI) by bootstrapping
  5. For the CI of the Hazard ratio, an additional method is available: inflate SE of hazard ratio based on the ITT test used to estimate λ

SE= standard error, CI = confidence interval

Hazard ratios for overall survival

Scenario	HR	95% CI (Bootstrap method for RPSFT)	95% CI (ITT method for RPSFT)
ITT	0.80	(0.68, 0.94)	
Censored	0.57	(0.48, 0.67)	
RPSFT	0.75	(0.65, 0.89)	(0.61, 0.92)

Overall survival
RPSFT-adjusted Kaplan-Meier curve



$\lambda = 1.43$

$\lambda^{-1} = 0.70$

Many thanks for our attention!
Any questions?

References for RPSFT

1. Robins JM, Tsiatis AA: Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics-Theory and Methods 1991, 20(8):2609-2631

2. Korhonen P, Zuber E, Branson M, Hollaender N, Yateman N, Katiskalahti T, Lebwohl D, 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

3. White IR, Walker S, Babiker AG, Darbyshire JH: Impact of treatment changes on the interpretation of the Concorde trial. Aids 1997, 11(8):999-1006

4. Branson M, Whitehead J: Estimating a treatment effect in survival studies in which patients switch treatment. Statistics in Medicine 2002, 21:2449-2463

5. http://www.analysisgroup.com/uploadedFiles/News_and_Events/News/ISPOR2011_Selected_Crossover_Cost-Effectiveness_Cancer_Therapies.pdf

6. Efron B, Tibshirini R: An introduction to the bootstrap. Monographs on Statistics and Applied Probability. Chapman and Hall London 1993.

7. White IR, Walker S, Babiker A: strbee: Randomization-based efficacy estimator. The Stata Journal 2002, 2(Number 2):140-150

Back-ups

Handling Censored Data

Diagram from Korhonen et al

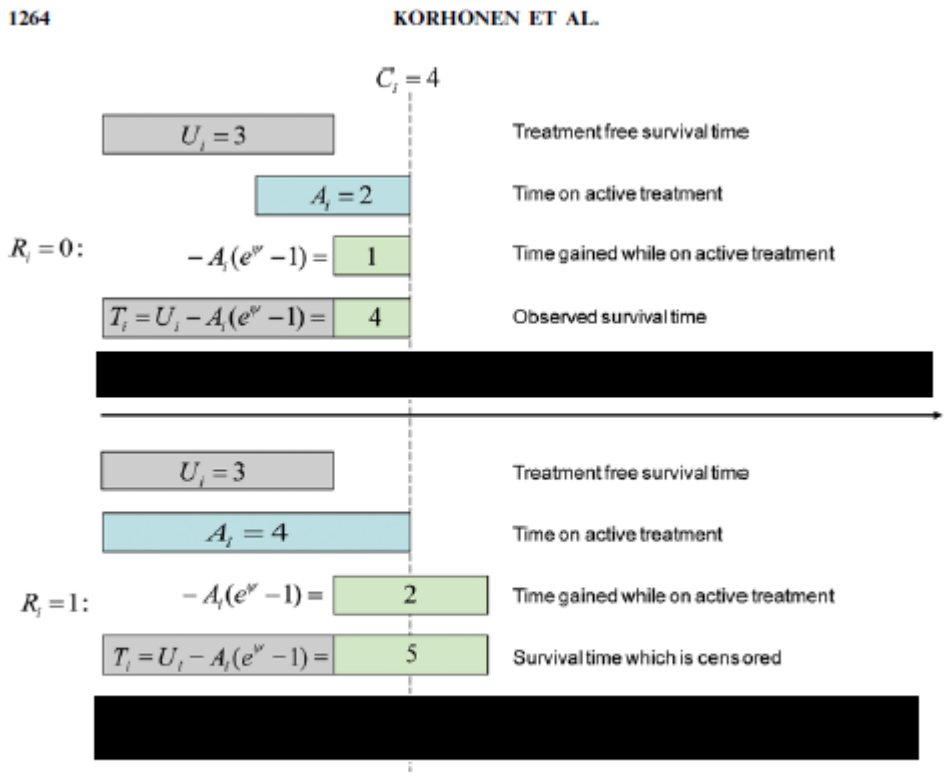


Figure 1 Illustrative example of calculation of underlying quantities in estimation procedure for patient randomized to placebo ($R_i = 0$) or everolimus ($R_i = 1$). (Color figure available online.)

Handling Censored Data (cont)

Diagram from Korhonen et al

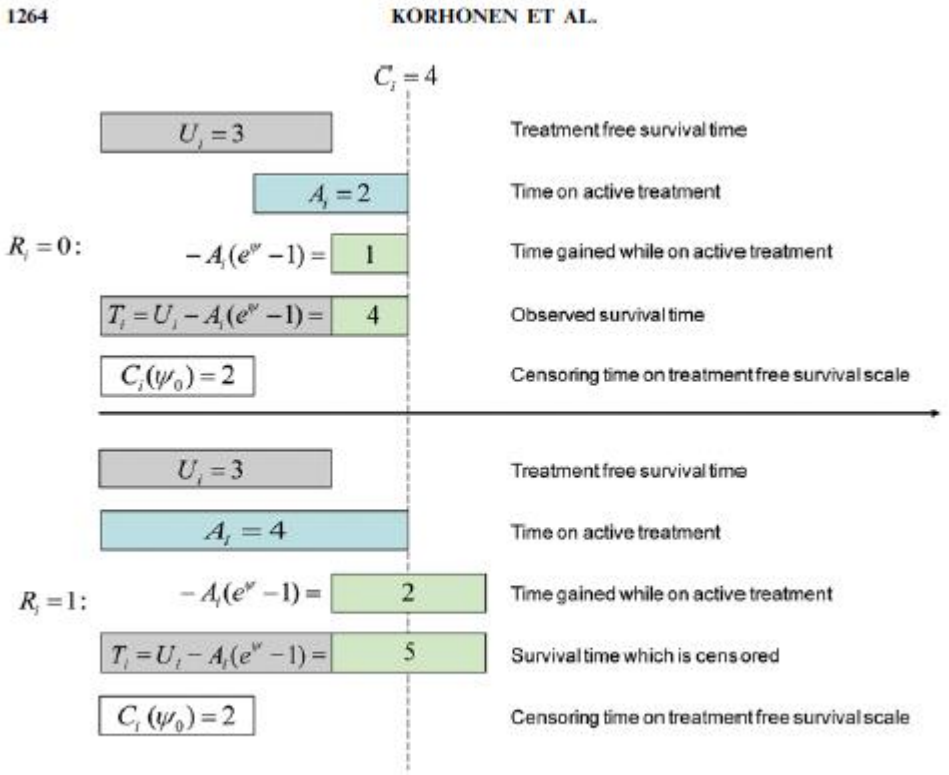


Figure 1 Illustrative example of calculation of underlying quantities in estimation procedure for patient randomized to placebo ($R_i = 0$) or everolimus ($R_i = 1$). (Color figure available online.)

Handling Censored Data (cont)

- Problem: uninformative administrative censoring at time C_i on T time scale is no longer uninformative on the U time scale
- Consider 2 patients A and B with latent survival time $U_i = 3$ months, and administrative censoring time $C_i = 4$ months with an acceleration factor $\lambda = 2$
- Patient A is randomized to control and crosses over at time $x_i = 2$ so is exposed to active treatment for 2 months and has a survival time of T_i = 4 months (3 months + 1 month extra) which will be observed
- Patient B is randomized to active so is exposed to active treatment from $x_i = 0$ so 4 months and would have a survival time T_i = 5 months (3 months + 2 months extra) which will be censored so we observe T^*_i = 4
- So, for data with censoring we can not just define $U^*_i = T_{Pi}^* + \lambda^{-1} T_{Ei}^*$ and $U_i = T_{Pi} + \lambda^{-1} T_{Ei}$ as this would be dependent on treatment history ($U_i = 3$ = 2+1 months for Patient A and $U^*_i = 2$ = 0+2 months for Patient B)

Adjusted CI for HR from Cox regression: ITT method (Inflation of standard error based on ITT test)

- Get Z_{ITT} (standard score of the test used to estimate λ) when comparing the T_i as randomized

- Define $SE_{ITT} = \frac{\log(HR)}{Z_{ITT}}$

- Define 95% CI in usual way as $\exp(\log(HR) \pm 1.96\, SE_{ITT})$

Seems to be comparable to bootstrap results

Adjusted CI for HR from Cox regression: The bootstrap method

- Generate samples (with replacement within treatment group/strata)
- Estimate λ for each sample
- Apply each samples λ to derive corrected survival times for that sample
- Estimate HR from comparing sample derived U_i and T_i
- Discard lowest 2.5% and highest 2.5% of bootstrap HRs to get 95% CI for HR

Open Topics

- In literature simple models tend to be used. Is this just because easier to estimate parameters? Or are these the best model?
- Assumptions require model to be correct – can this be checked? How sensitive are the results to different models relating U_i to T_i?
- Should we fit multiple models in a parametric fashion – then select best fitting? Investigate the Hybrid approach?
- Should we define our model before unblinding if we expect crossover to be allowed later?
- In practice is there any benefit in using optimum/near optimum weights in the test used for equality of the latent survival time? Can these even be defined?
- Is it possible to define a less aggressive re-censoring algorithm – can we accept some bias in a trade-off against losing information?
- How can we capture uncertainty in other estimates without bootstrapping?