U.S. FOOD & DRUG ADMINISTRATION

# Comparison of Hazard Ratio and Restricted Mean Survival Analysis for Cardiorenal Drug Trials

**John Lawrence, Junshan Qiu,**

**Steven Bai, Jim Hung**

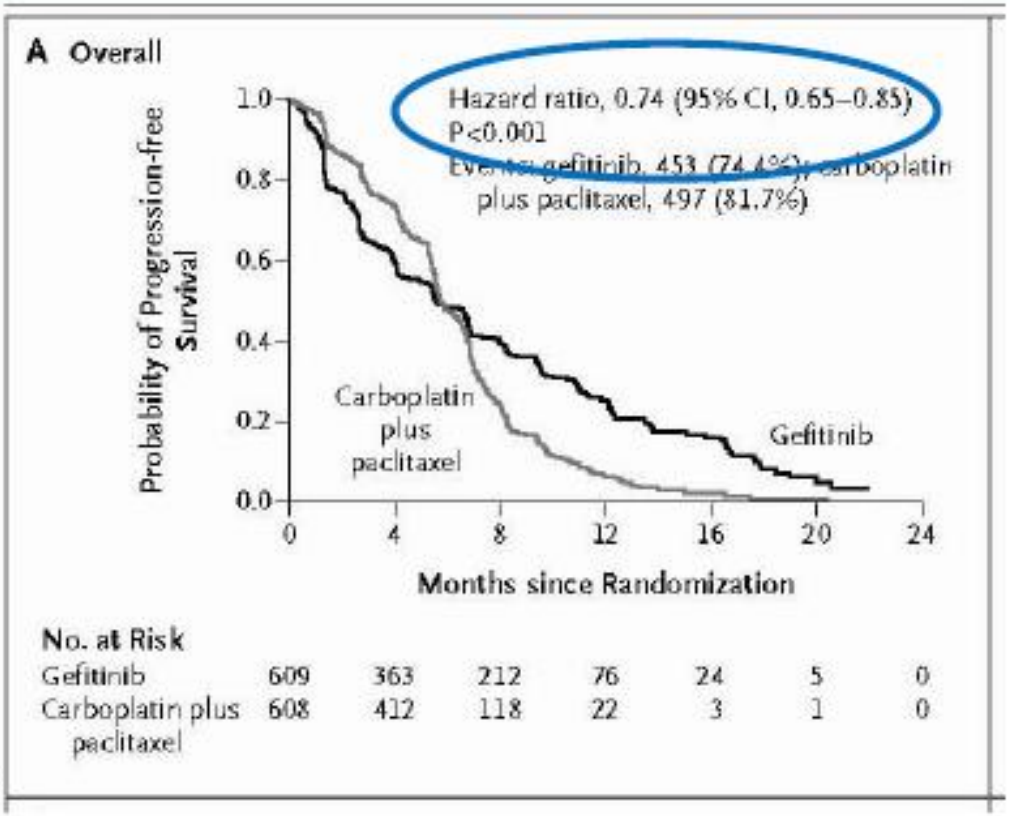
# Disclaimer

This presentation reflects the views of the authors and should not be construed to represent FDA's views or policies.

# The Problem with Hazard Ratio

- Hazard Ratio may not be constant
- Difficult to interpret, even if it is constant, but even more so if it is not constant

# … non-ignorable non-PH does happen!




From slide presentation by Royston and Parmar
https://statmarker.sciencesconf.org/data/pages/Parmar_Presentation_STATMAKER_20_10_2016.pdf

## What does the overall hazard ratio mean?

- In the reconstructed IPASS example, the HR ranges between 0.27 and 2.2 over time
- The overall HR at the time of this analysis is 0.73 (95% CI 0.64, 0.83)
- What does this mean?
- Some people (e.g. Schemper 2009) have interpreted the overall HR as a type of a weighted average HR over the event times
- But we think a single HR when there is non-PH is not interpretable
- Instead we work with RMST



From slide presentation by Royston and Parmar
https://statmarker.sciencesconf.org/data/pages/Parmar_Presentation_STATMAKER_20_10_2016.pdf

# Background

- To estimate treatment effect for time to event, Hazard Ratio (HR) is commonly used.
- HR is often assumed to be constant over time (i.e., proportional hazard assumption).
- Recently, we have some doubt about this assumption.
- If the PH assumption does not hold, the interpretation of HR can be difficult.

# Background

- Another measure of treatment effect could be based on median, but in the CV trials, median survival time is hardly calculable due to small event rates.

- Rather than the median (the $50^{th}$ percentile), another option could be a different quantile, e.g. the $90^{th}$ percentile.

- In one group, 90% of the people survive at least x days, in the other group 90% of the people survive at least y days.

- Much information could be lost because the actual survival times (for those greater than the $90^{th}$ percentile) are not used.

# Background

- What about comparing mean survival time?

- In the time-to-event analysis, if the last observation is censored, mean cannot be estimated from Kaplan-Meier curve without making some assumptions about the distribution beyond the last event time.

- Survival data is often skewed to the right and in some situations the median is preferred over the mean as summary measure.

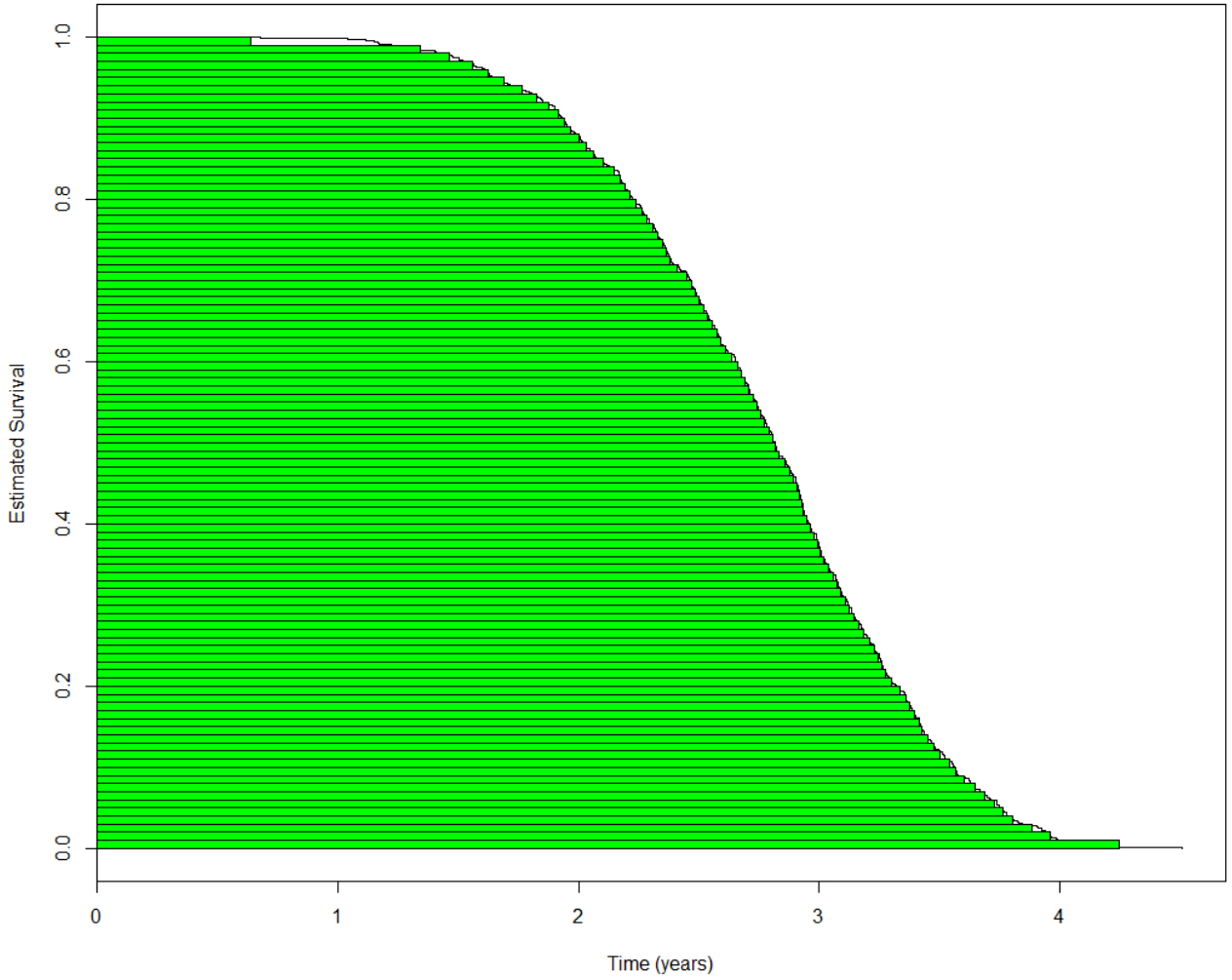
# Background

- The idea of Restricted Mean Survival Time (RMST) goes back to Irwin (1949) and is further implemented in survival analysis by Uno et al. (2014).

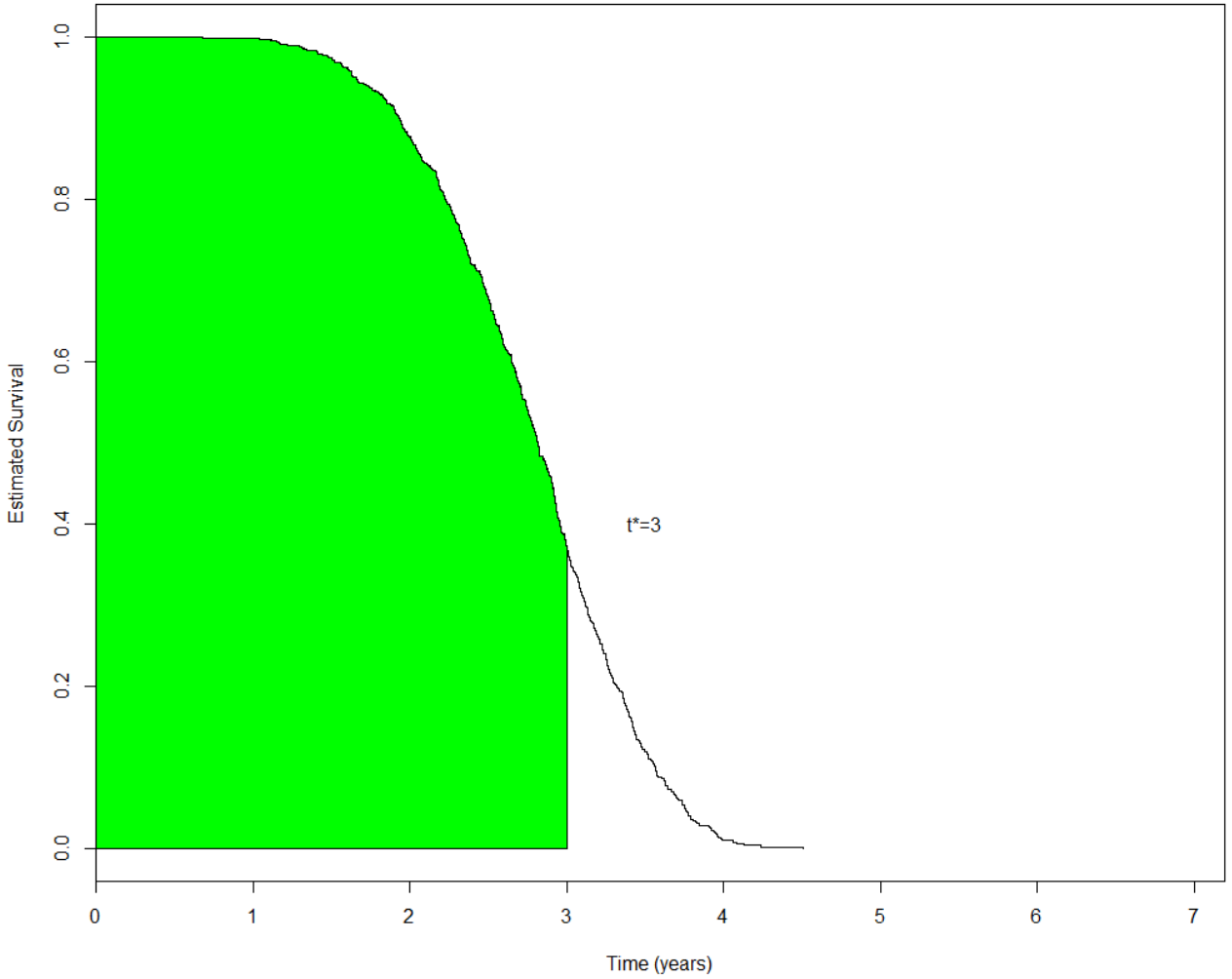
- RMST is defined as the area under the survival curve up to t*, which should be pre-specified for a randomized trial.

- RMST for a CV outcome may be loosely described as the event free expectancy over the restricted period between randomization and a defined, clinically relevant time horizon, called t*.

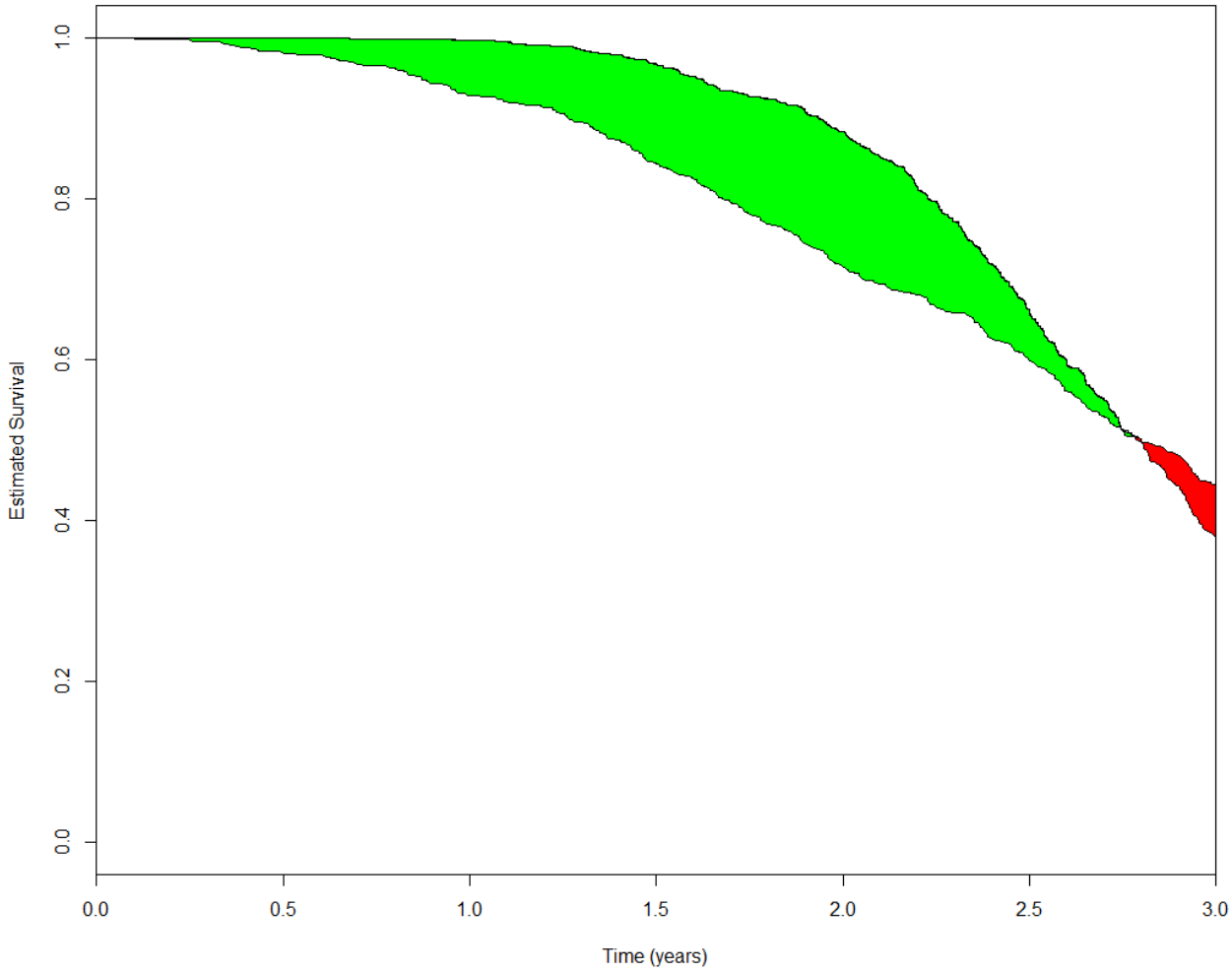
# Mean as Area Under Curve



# RMST as Area Under Curve



# Difference in RMST



# RMST- Small Sample Distribution

- We observed that when testing for non-inferiority with margin $M$ and a small number of events (<50), the test statistic

$$\frac{\hat{\mu}_1 - \hat{\mu}_0 + M}{\sqrt{\widehat{Var}\{\hat{\mu}_1 - \hat{\mu}_0\}}}$$

  is not very well approximated by N(0,1).

- We don't think the approximation is poor when testing for superiority because theoretically the ratio should have mean 0 and be symmetric.

# RMST- Small Sample, NI testing

- Simulations under Example Scenario from SSRMST package
  uniform accrual over 35 days
  Total study time=510 days
  truncation time = 500 days
  Margin =18 days
  Exp survival times with parameters to make difference in RMST = 18 days
  500,000 runs

| Central Moment | Observed Sample Value |
|---|---|
| Mean | 0.0284 |
| Variance | 1.015 |
| Third | 0.119 |
| Fourth | 3.09 |
| Fifth | 1.179 |

# RMST- Small Sample, NI testing

- Although $\hat{\mu}_1 - \hat{\mu}_0 + M$ has mean 0, the numerator and denominator of

$$\frac{\hat{\mu}_1 - \hat{\mu}_0 + M}{\sqrt{\widehat{Var}\{\hat{\mu}_1 - \hat{\mu}_0\}}}$$

  are not independent, so the ratio does not have mean 0.

# RMST- Small Sample Distribution

- Ad hoc correction: Find expected value of ratio, standardize the statistic to have mean 0 and use a t-distribution with d.f. found by Satterthwaite approximation using number of events (not number of subjects).

# RMST- Small Sample Distribution

- Better correction:

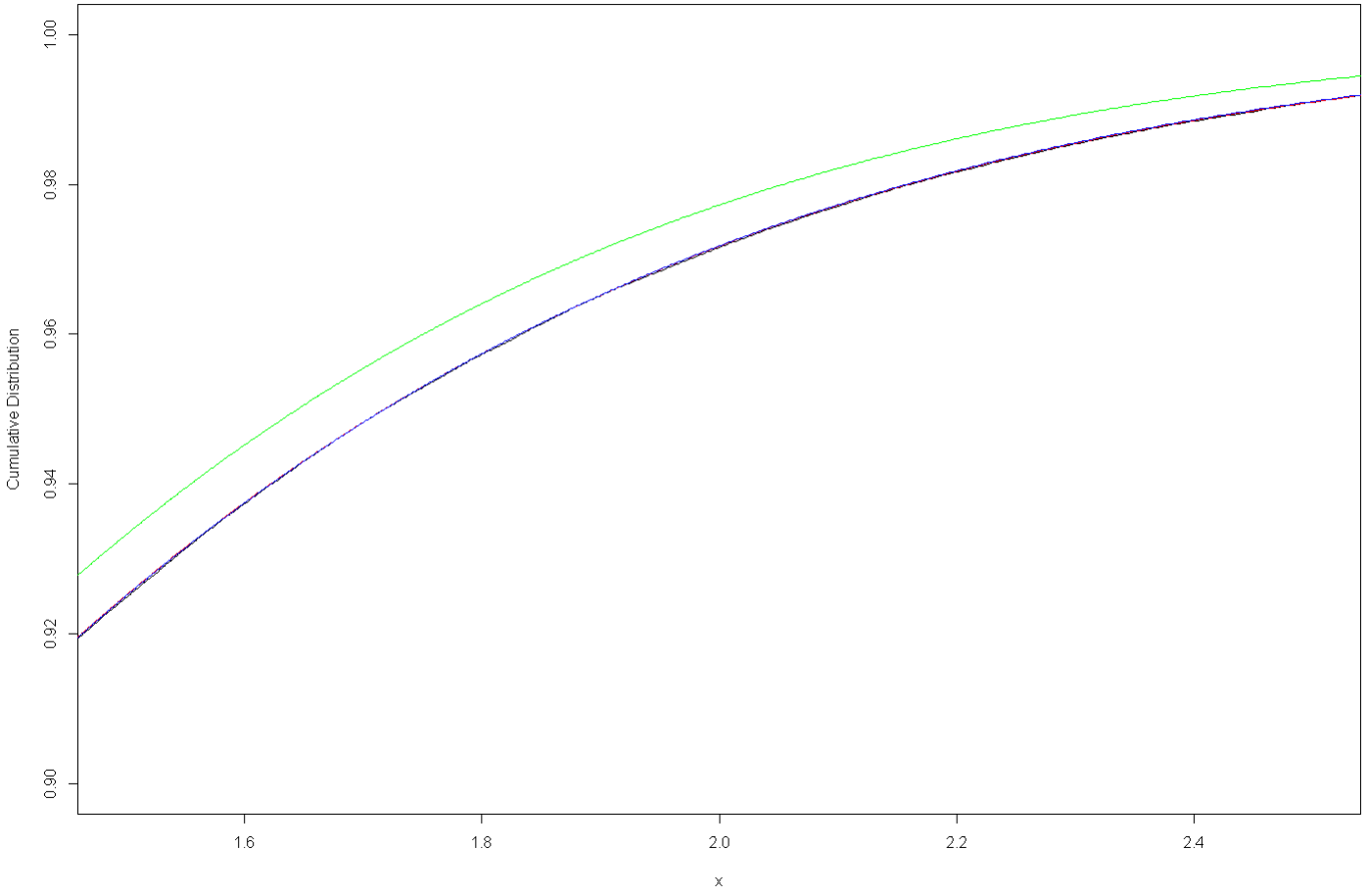
  Estimate cumulants of the statistic

  Use Cornish-Fisher expansion

  This correction can go either way relative to the normal approximation- it does not always make a correction that reduces the power

# Cornish Fisher Expansion

(Green=Normal distribution. Black, red, blue = Empirical CDF and Cornish-Fisher expansions of different orders



# RMST vs. HR Asymptotic Efficiency

Asymptotic mean and variance of logrank statistic (Schoenfeld, D. 1981. ).

For RMST, mean can be found.

$$N\widehat{Var}\{\hat{\mu}\} \rightarrow \int_0^{t^*} \left[ \int_x^{t^*} S(t)dt \right]^2 \frac{f(x)}{S(x)^2\{1-H(x)\}} dx$$

Use this to find sample size for any given censoring distributions and survival curves.
A.R.E. is square of ratio of slopes

# RMST vs. HR Asymptotic Efficiency

Assume Weibull Distributions

Hazard function is $\frac{k\, x^{k-1}}{\lambda^{k}}$

with scale parameter $\lambda$ and shape parameter $k$.

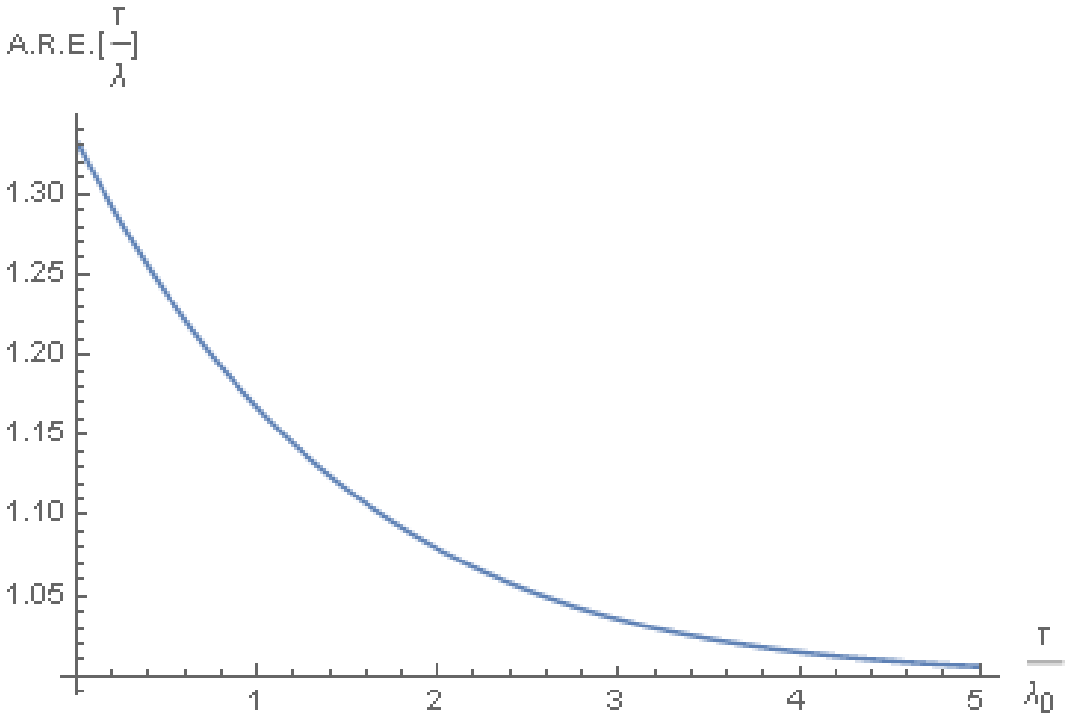
# RMST vs. HR Asymptotic Efficiency

Scenarios

1) Same shape parameters. Scale parameters are $\lambda_0$ and $\lambda_1$ with $\lambda_1 \to \lambda_0$. Hazard ratio is $\left(\frac{\lambda_0}{\lambda_1}\right)^k$

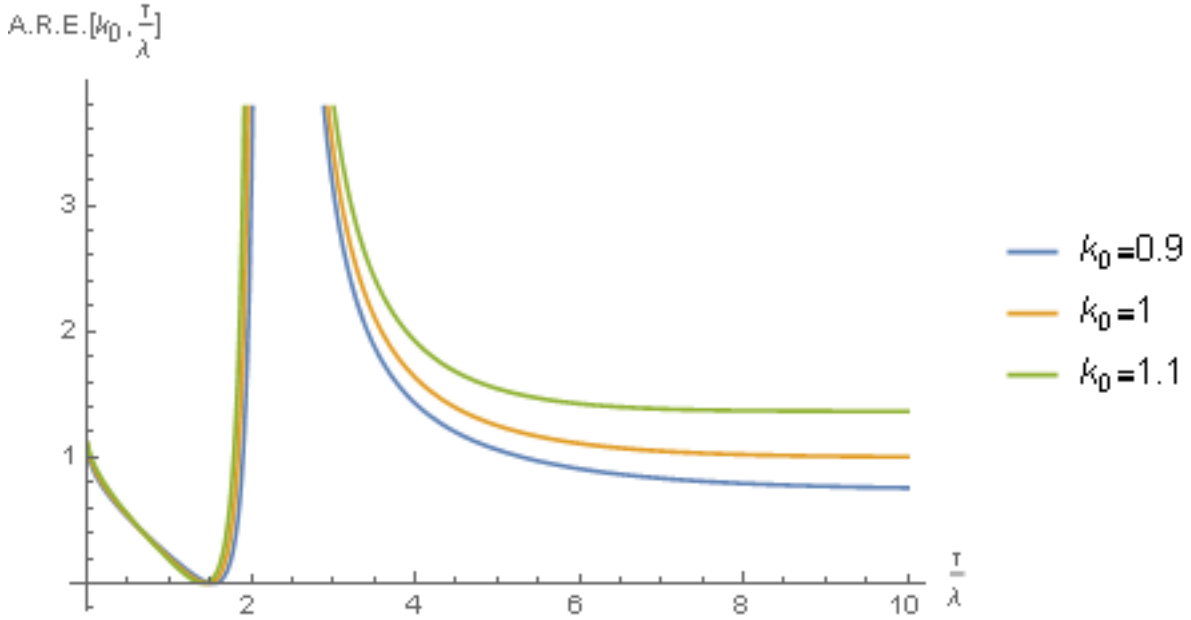
2) Same scale parameters. Shape parameters are $k_0$ and $k_1$ with $k_1 \to k_0$

# A.R.E. Scenario 1




Not surprisingly, the logrank test is more efficient regardless of the truncation time $\tau$.
If $k$=1 and $\tau$=$\lambda_0$, you need about 17% more patients if using RMST compared to logrank test.

# A.R.E. Scenario 2




For some $\tau$, RMST is more efficient.
This figure is for two-sided testing.

For some $\tau$, the mean is in the opposite directions (RMST has power to show drug is harmful, but logrank test has power to show drug is beneficial).

For true one-sided testing, neither test has power for $\frac{\tau}{\lambda} < 1.5$, logrank would be infinitely more powerful between 1.5 and about 3.

# Power and ARE

For some alternatives and some $\tau$, RMST is more efficient.

RMST is not being recommended because it is more powerful or reduces the sample size, it is because it is easier to interpret.

# Power and ARE

If increased power is desired and hazards are not proportional, change the weights!

From Schoenfeld 1981

5. CONSTRUCTING OPTIMAL TESTS

To find the most powerful test that can be put in form (1) we note that, as a consequence of the Cauchy–Schwarz inequality, (2) is maximized when $g(t)$ is proportional to $\log\{\lambda_1(t)/\lambda_0(t)\}$. This is reasonable as we would want to put large weights on $X_j - p(t_j)$ when the hazard on treatment 0 is much greater than that of treatment 1; we are rejecting

Weights can be chosen using the data at an interim analysis. This can be done in a way that controls the type 1 error rate and this strategy can increase the power significantly. Lawrence, J. 2002.

# Background

- The primary endpoint for CV outcome trials is usually a composite endpoint comprised of major adverse cardiac events.
- These CV outcome trials are featured with
  - Low incidence rate of CV events
  - Large sample size
  - Long follow-up period

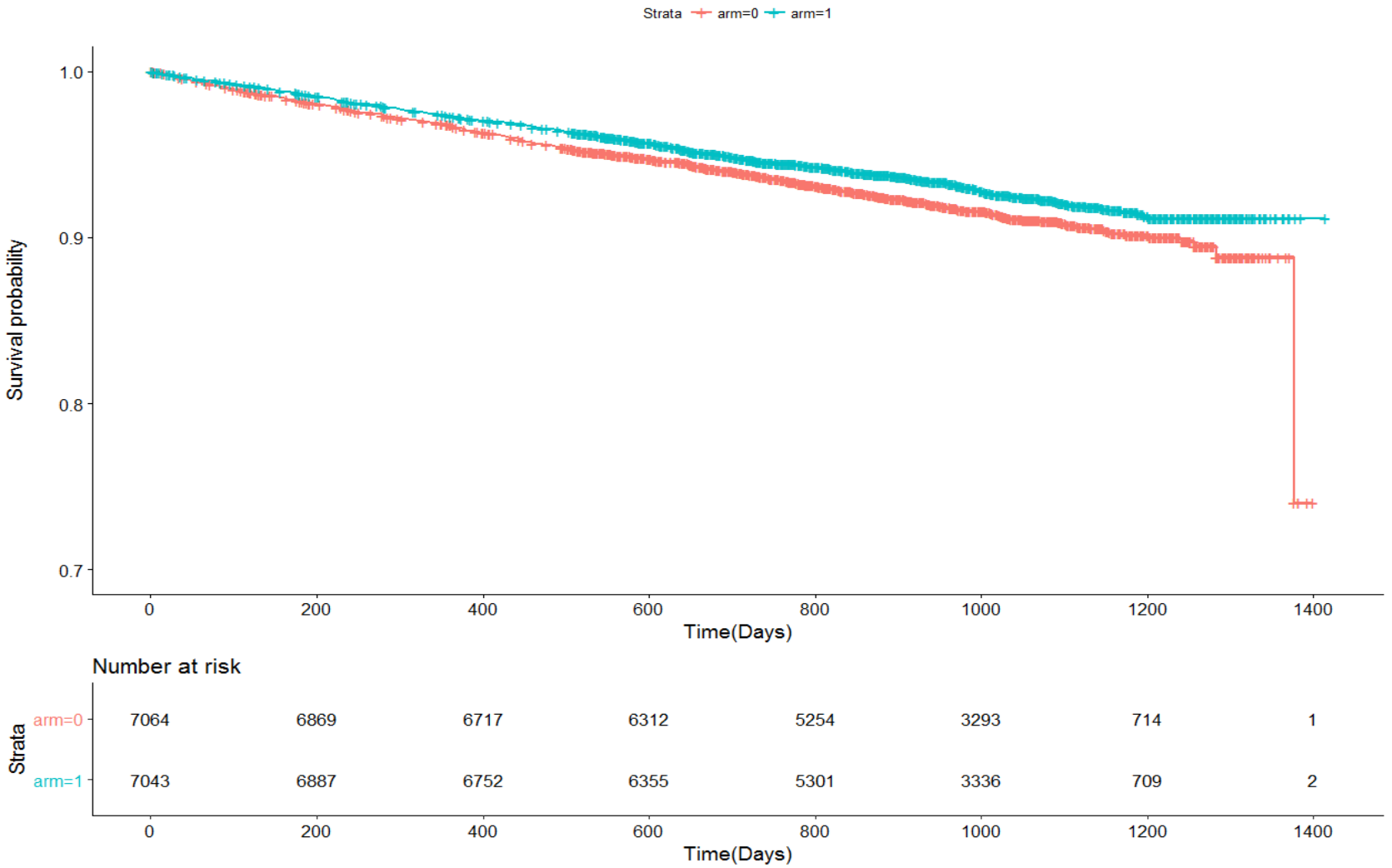
# Questions

- Should RMST be used as sensitivity analysis (when PH assumption is not met)?
- Can RMST be considered as the primary analysis?

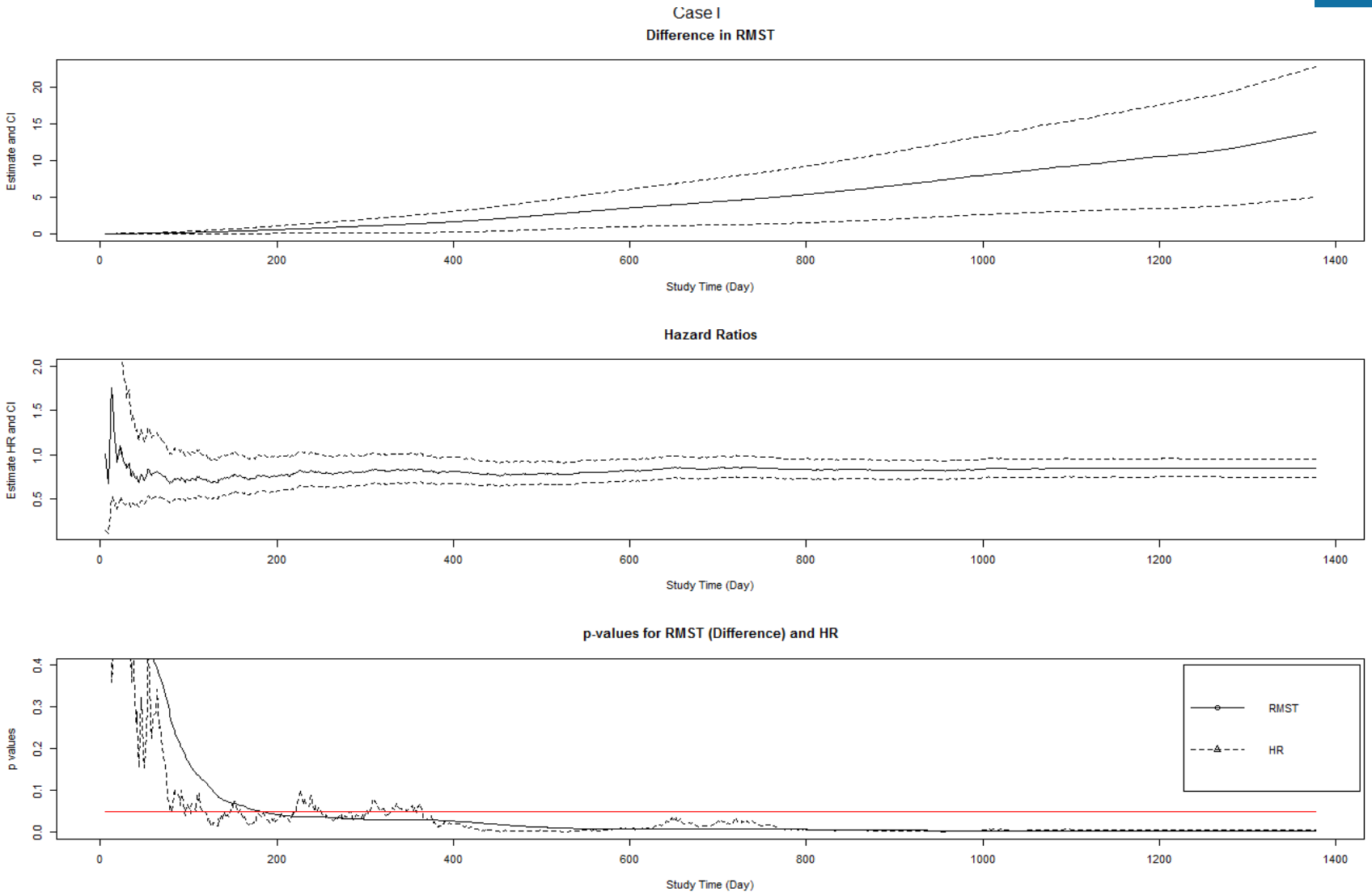
# Piloting with 6 Cases

- 6 cardiovascular clinical trials
- Low event rates
- Survival probability at the end of the trial > 50% except for Case VI
- Some cases may have non-proportional hazards due to delayed treatment effects, crossed survival curves (e.g., unstable treatment effects), or diluted treatment effects.

# Case I


Strata arm=0 arm=1
1.0
0.9
0.8
0.7
Survival probability
0
200
400
600
800
1000
1200
1400
Time(Days)
Number at risk
Strata
arm=0
arm=1
7064
6869
6717
6312
5254
3293
714
1
7043
6887
6752
6355
5301
3336
709
2
0
200
400
600
800
1000
1200
1400
Time(Days)

# Case I

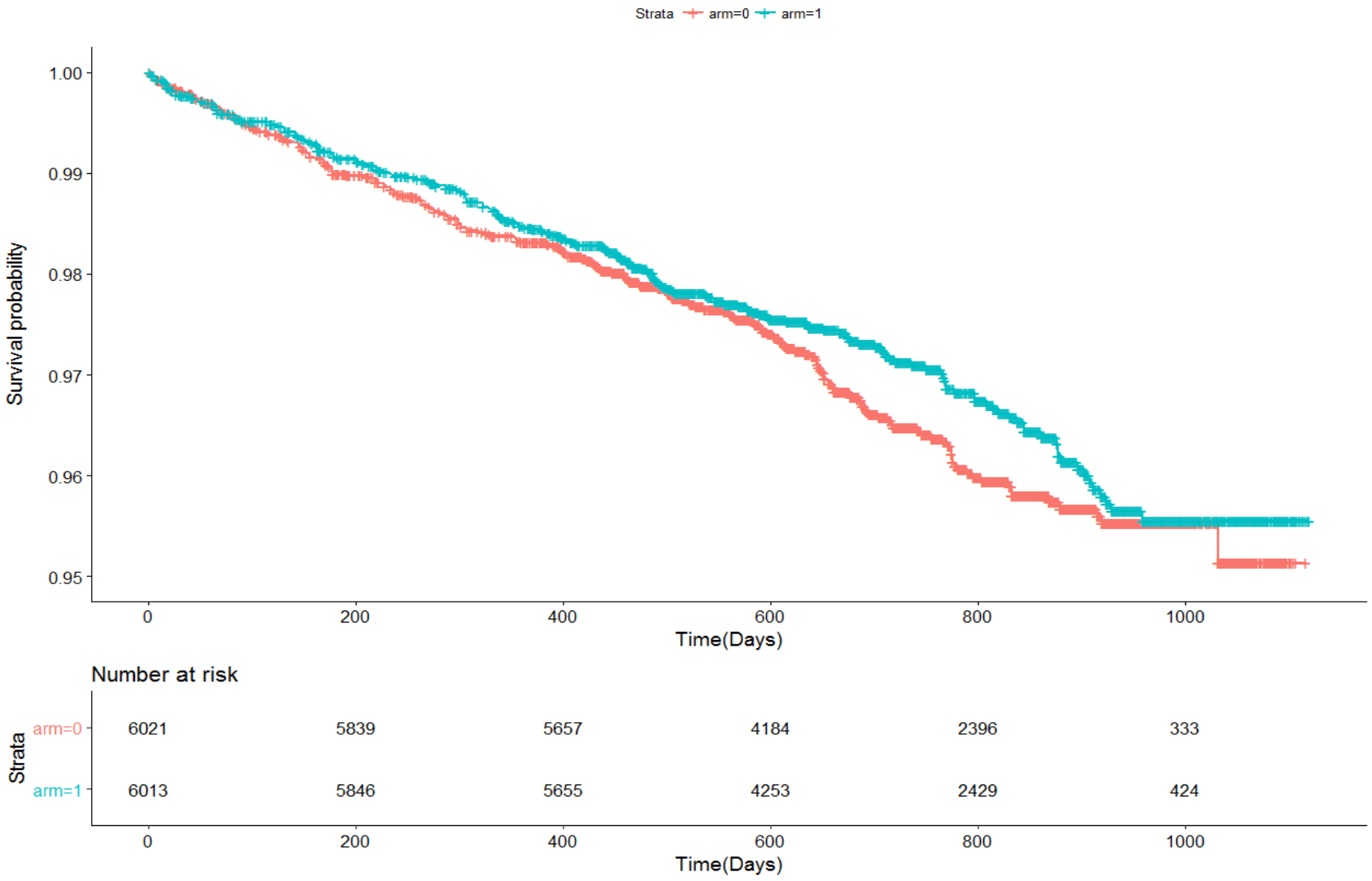

Case I
Difference in RMST
Estimate and CI
0
5
10
15
20
0
200
400
600
800
1000
1200
1400
Study Time (Day)
Hazard Ratios
Estimate HR and CI
0.5
1.0
1.5
2.0
0
200
400
600
800
1000
1200
1400
Study Time (Day)
p-values for RMST (Difference) and HR
p values
0.0
0.1
0.2
0.3
0.4
RMST
HR
0
200
400
600
800
1000
1200
1400
Study Time (Day)

# Case I

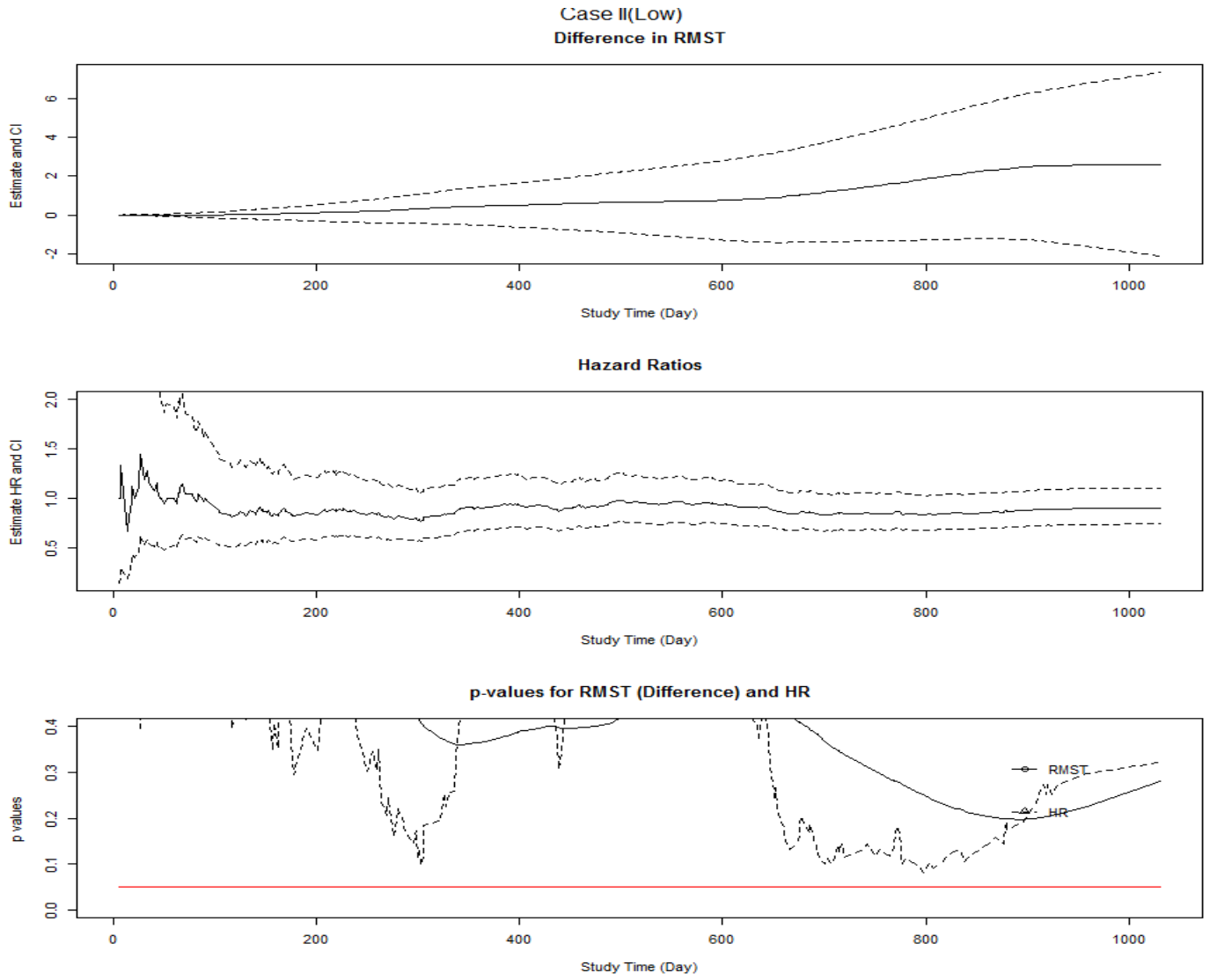
| | Estimate | 95% Lower | 95% Upper | p value |
|---|---|---|---|---|
| Diff in RMST | 13.84 | 4.93 | 22.75 | 0.0023 |
| Cox Reg HR | 0.84 | 0.75 | 0.95 | 0.0048 |

The RMST method yields a smaller p-value than the Cox reg HR. The first occurrence of any composite event can be delayed by 13.8 days on average over 4 years for the patients who are on the Drug A as compared to the control.

# Case II


Strata arm=0 arm=1
1.00
0.99
0.98
0.97
0.96
0.95
Survival probability
0
200
400
600
800
1000
Time(Days)
Number at risk
Strata
arm=0
arm=1
6021
5839
5657
4184
2396
333
6013
5846
5655
4253
2429
424
0
200
400
600
800
1000
Time(Days)

# Case II


Case II(Low)
Difference in RMST
Estimate and CI
Study Time (Day)
Hazard Ratios
Estimate HR and CI
Study Time (Day)
p-values for RMST (Difference) and HR
p values
RMST
HR
Study Time (Day)

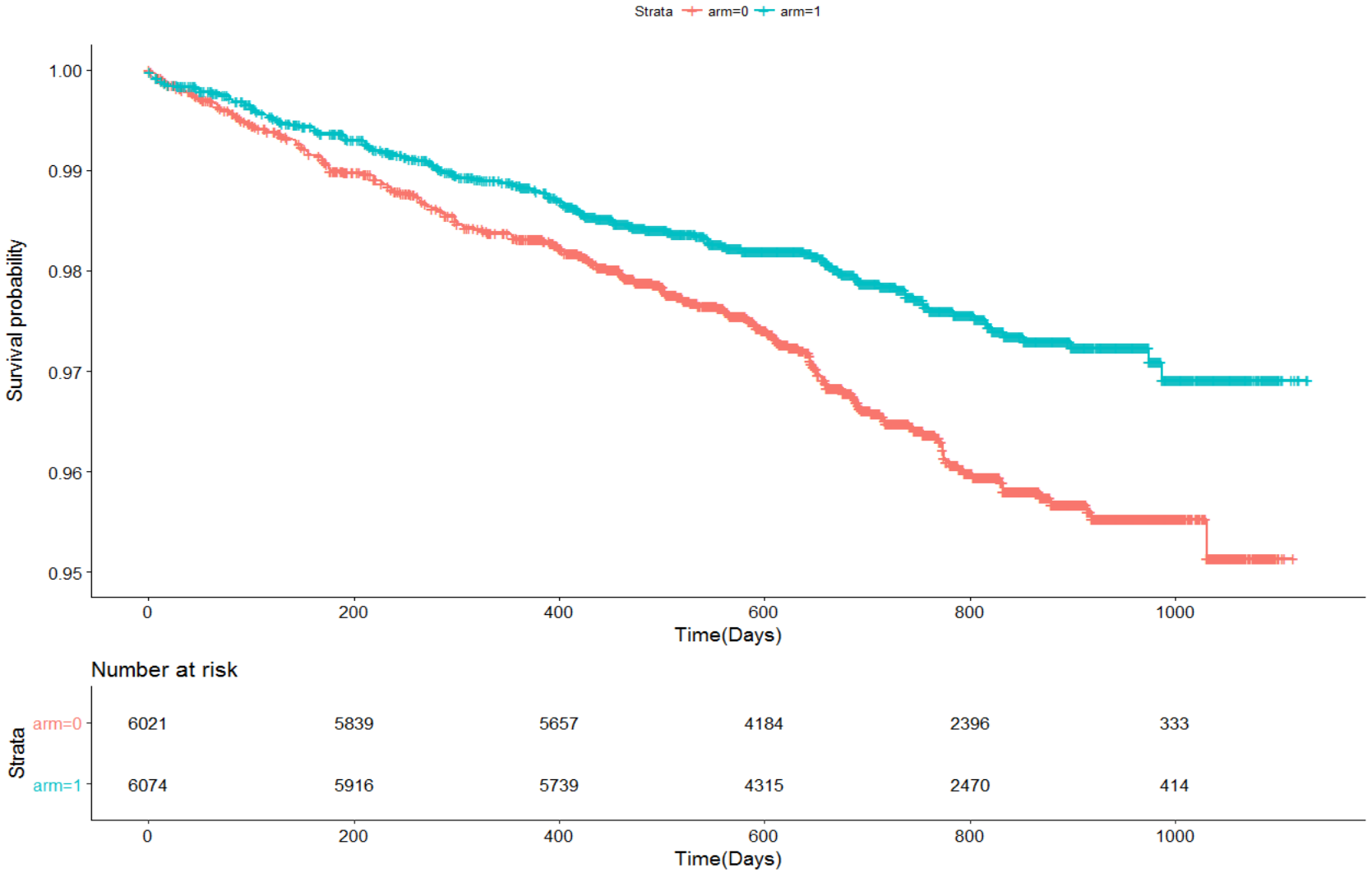
# Case II

Low

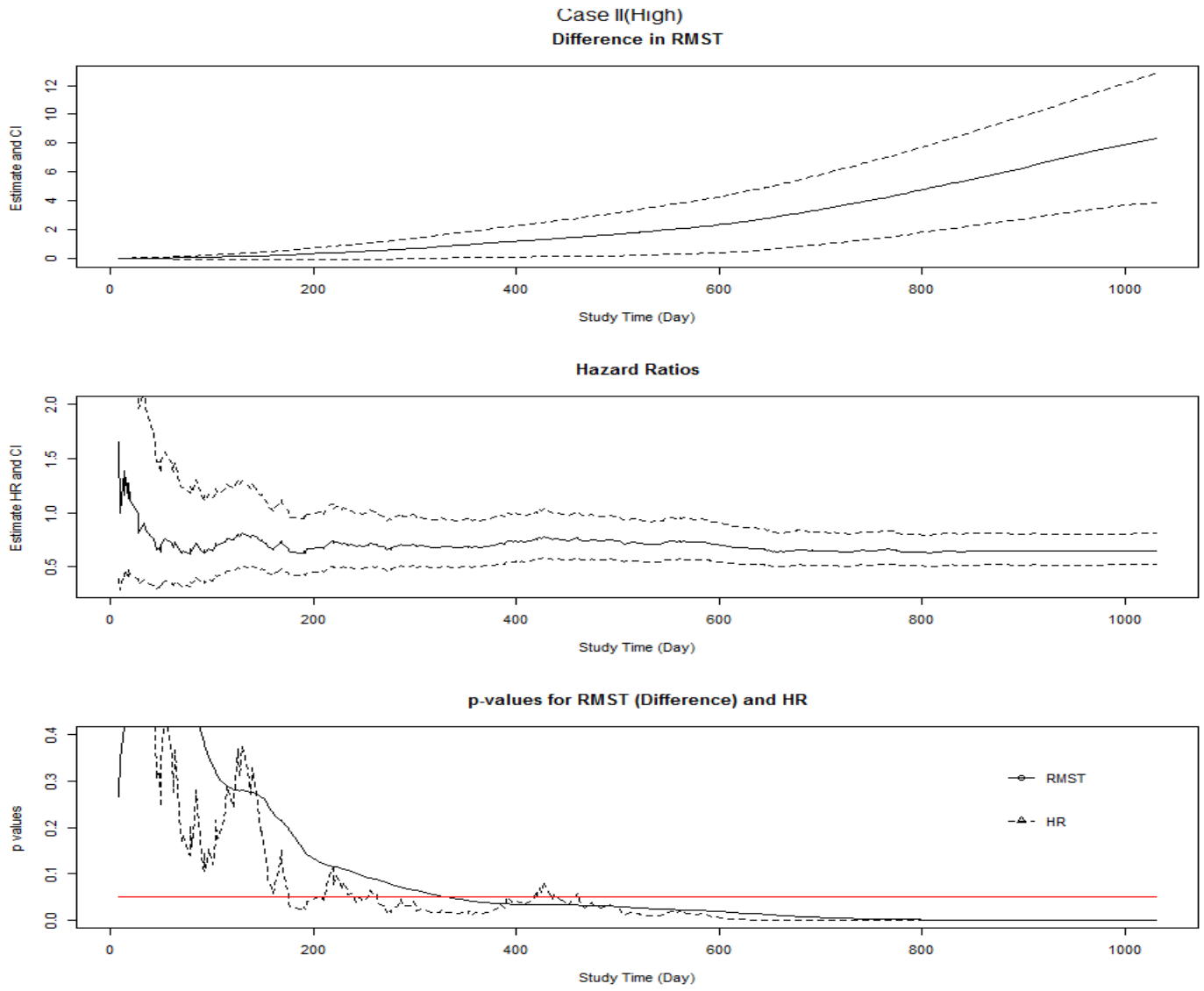
| | Estimate | 95% Lower | 95% Upper | p value |
|---|---|---|---|---|
| Diff in RMST | 2.61 | -2.14 | 7.35 | 0.28 |
| Cox Reg HR | 0.90 | 0.74 | 1.10 | 0.32 |

The RMST method provides a slightly smaller p-value the Cox reg HR. The first occurrence of any composite event can be delayed by 2.6 days on average over 3.5 years for the patients who are on Drug B at low dose level from the patients who are on Warfarin.

# Case II


Strata arm=0 arm=1
Survival probability
1.00
0.99
0.98
0.97
0.96
0.95
0
200
400
600
800
1000
Time(Days)
Number at risk
Strata
arm=0
arm=1
6021
5839
5657
4184
2396
333
6074
5916
5739
4315
2470
414
0
200
400
600
800
1000
Time(Days)

# Case II


Case II(High)
Difference in RMST
Estimate and CI
0
2
4
6
8
10
12
0
200
400
600
800
1000
Study Time (Day)
Hazard Ratios
Estimate HR and CI
0.5
1.0
1.5
2.0
0
200
400
600
800
1000
Study Time (Day)
p-values for RMST (Difference) and HR
p values
0.0
0.1
0.2
0.3
0.4
RMST
HR
0
200
400
600
800
1000
Study Time (Day)

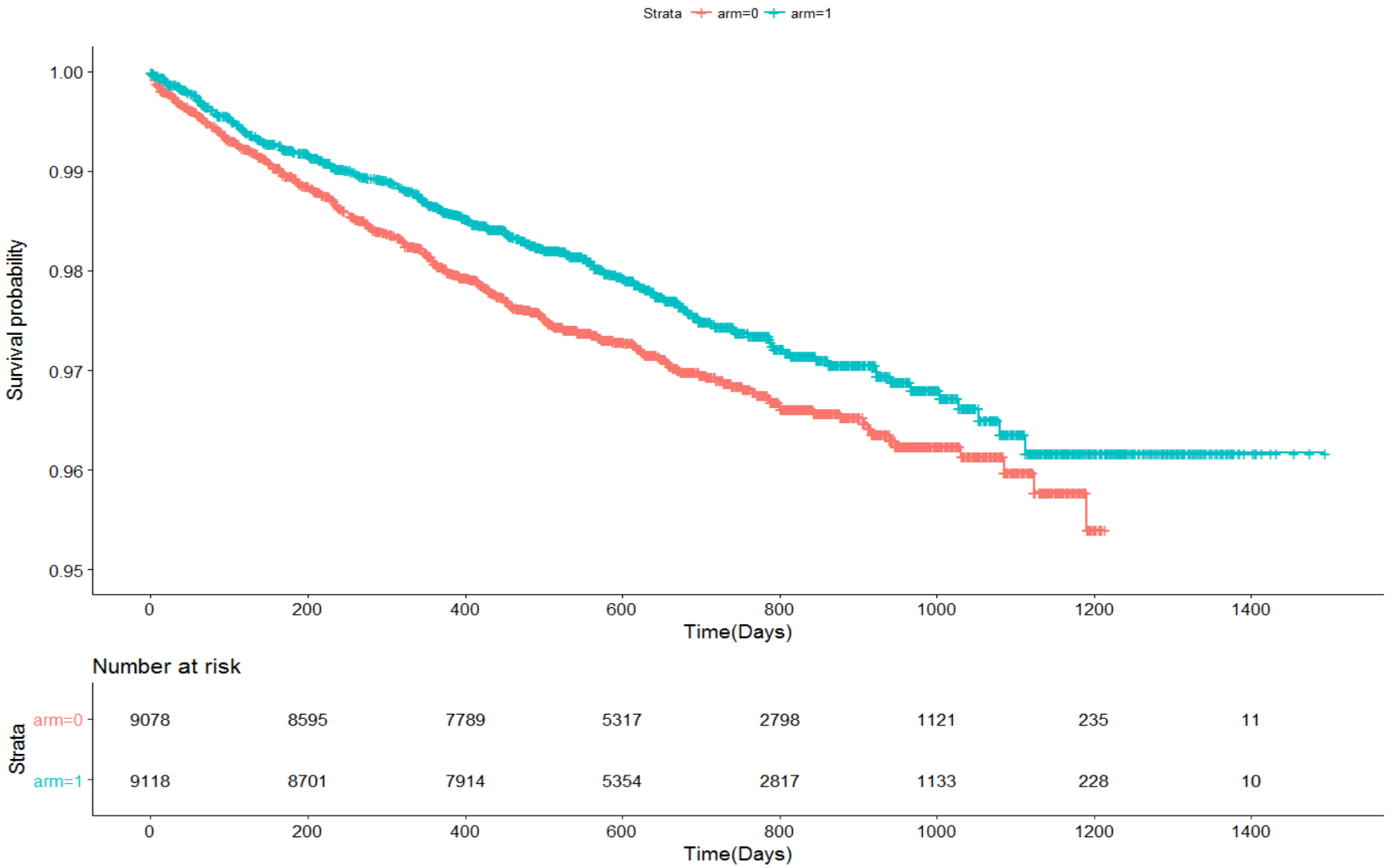
# Case II

High

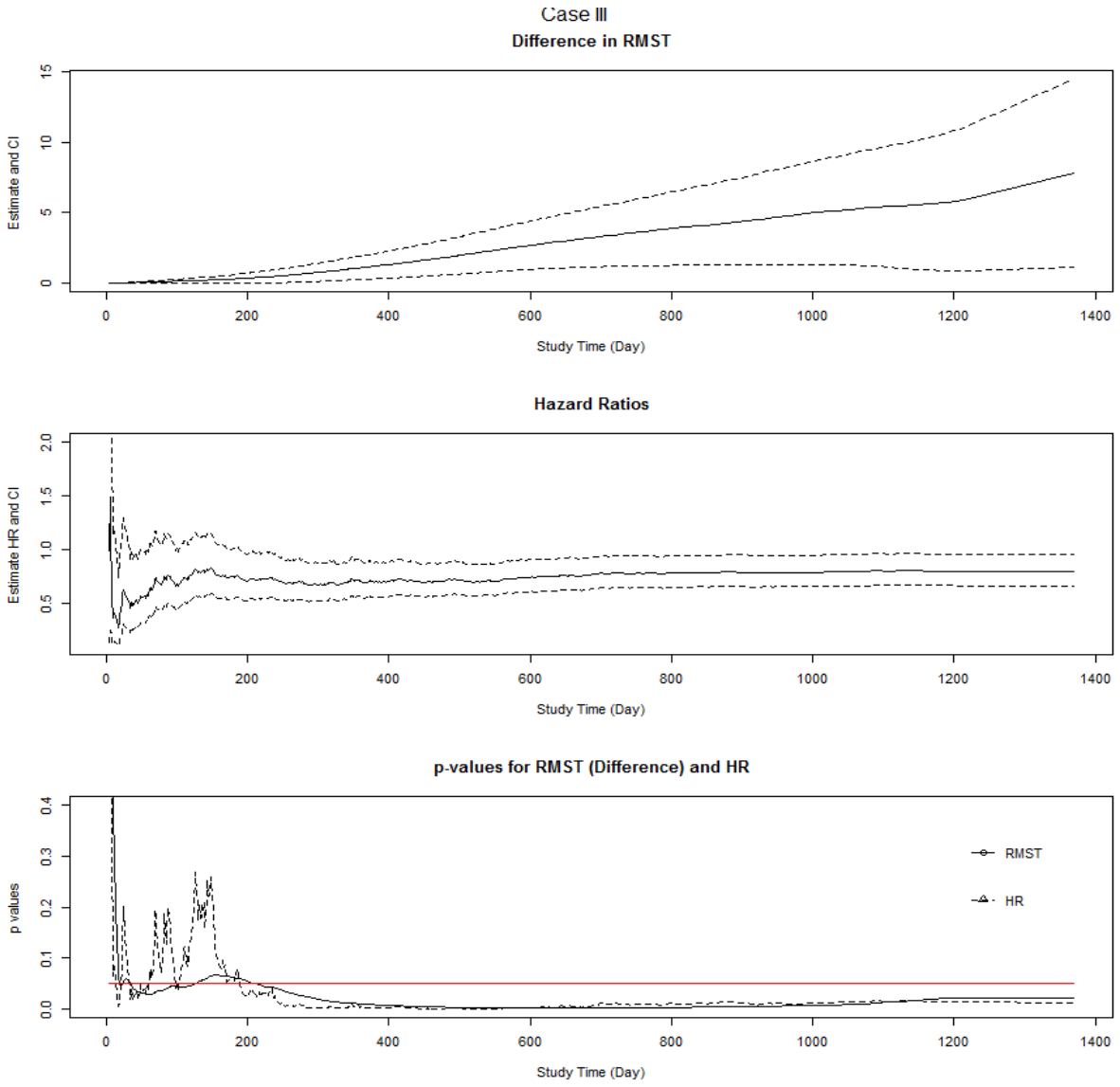
| | Estimate | 95% Lower | 95% Upper | p value |
|---|---|---|---|---|
| Diff in RMST | 8.34 | 3.88 | 12.81 | 2e-04 |
| Cox Reg HR | 0.65 | 0.52 | 0.81 | 1e-04 |

The RMST method provides a bigger p-value than the Cox reg HR. The first occurrence of any composite event can be delayed by 8.3 days on average over 3.5 years for the patients who are on the Drug B at the high dose level from the patients who are on Warfarin.

# Case III


Strata arm=0 arm=1
1.00
0.99
0.98
0.97
0.96
0.95
Survival probability
0
200
400
600
800
1000
1200
1400
Time(Days)
Number at risk
Strata
arm=0
arm=1
9078
8595
7789
5317
2798
1121
235
11
9118
8701
7914
5354
2817
1133
228
10
0
200
400
600
800
1000
1200
1400
Time(Days)

# Case III

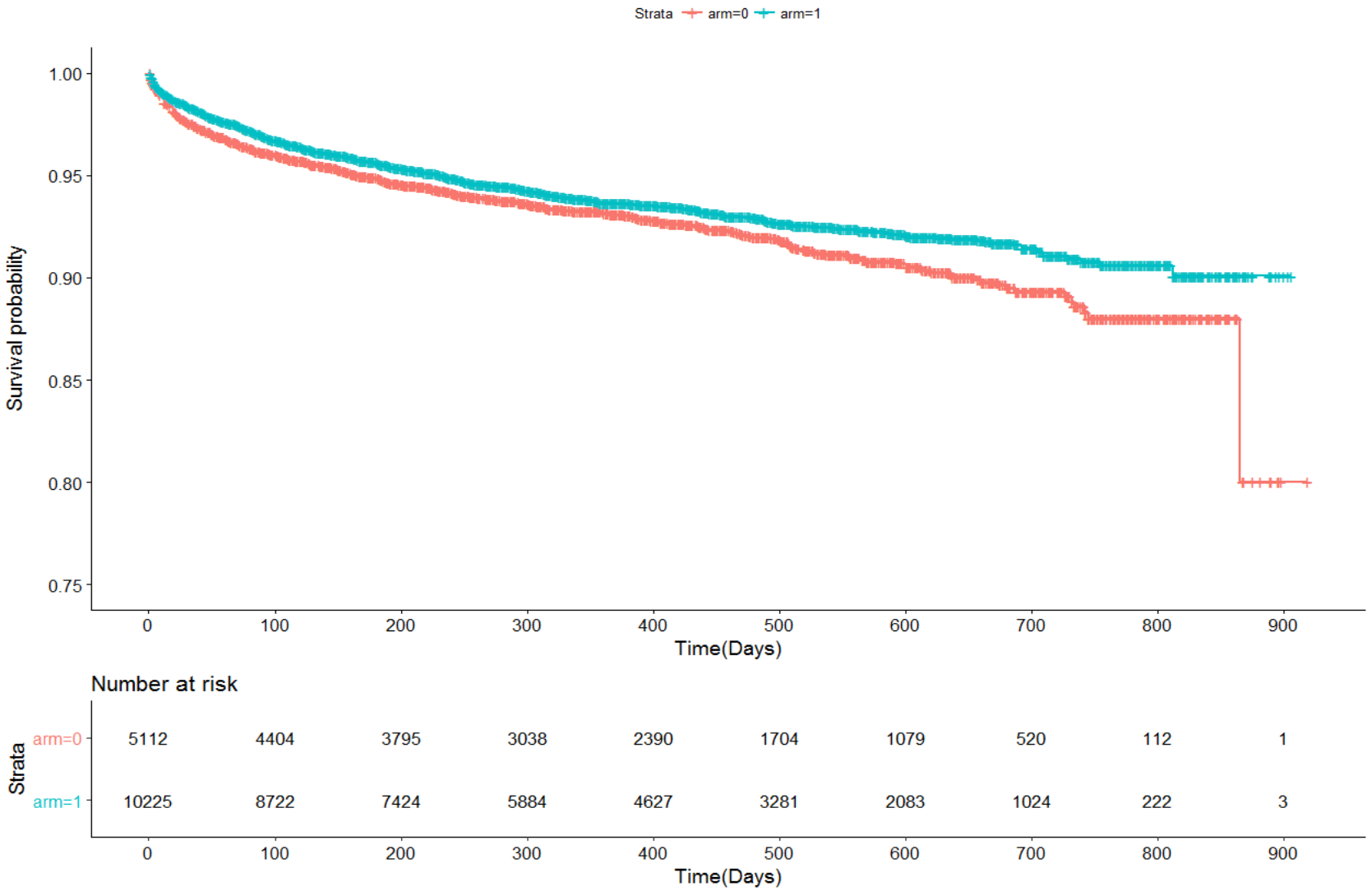

Case III
Difference in RMST
Estimate and CI
0
5
10
15
0
200
400
600
800
1000
1200
1400
Study Time (Day)
Hazard Ratios
Estimate HR and CI
0.5
1.0
1.5
2.0
0
200
400
600
800
1000
1200
1400
Study Time (Day)
p-values for RMST (Difference) and HR
p values
0.0
0.1
0.2
0.3
0.4
RMST
HR
0
200
400
600
800
1000
1200
1400
Study Time (Day)

# Case III

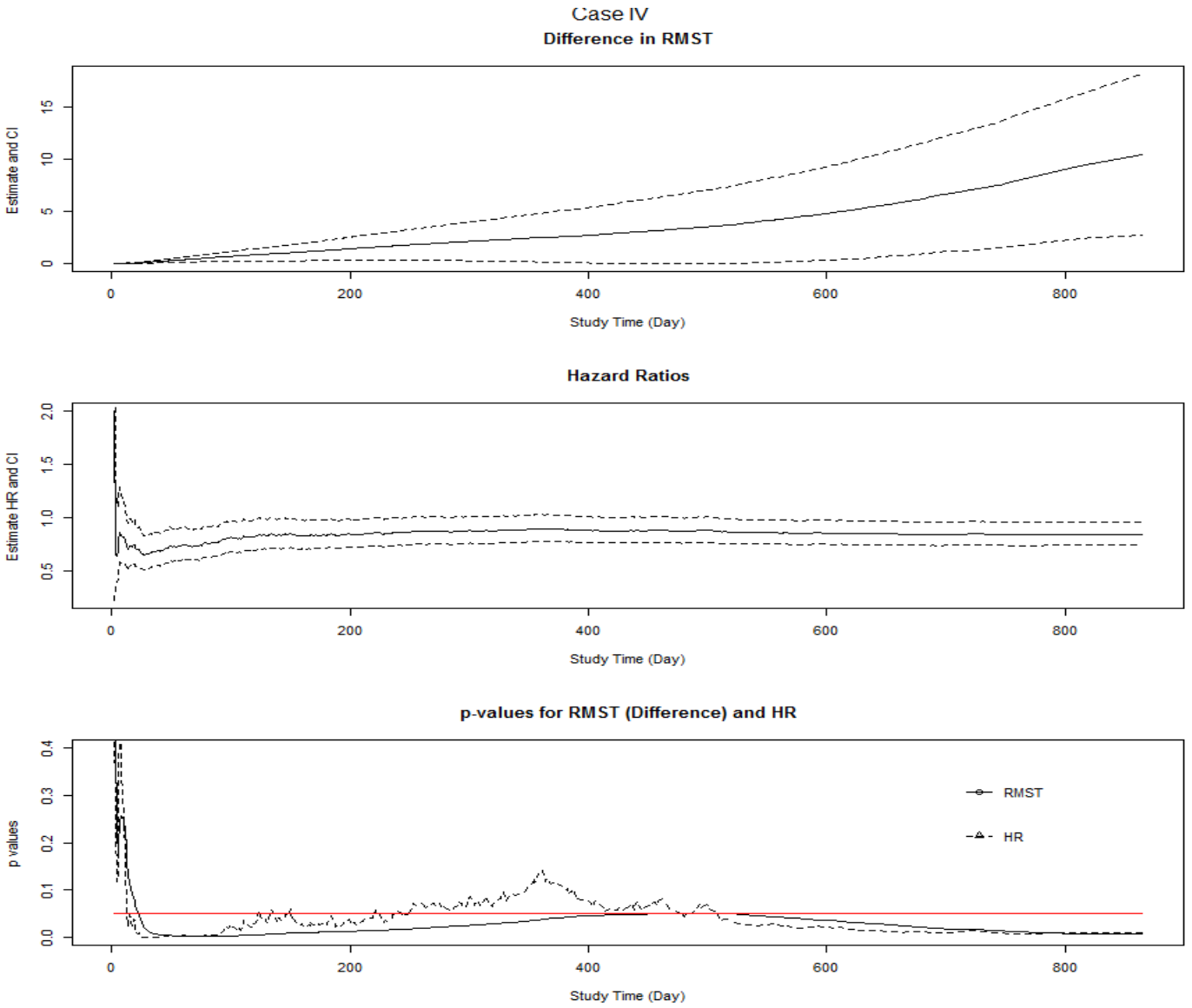
| | Estimate | 95% Lower | 95% Upper | p value |
|---|---|---|---|---|
| Diff in RMST | 7.81 | 1.15 | 14.46 | 0.022 |
| Cox Reg HR | 0.80 | 0.66 | 0.95 | 0.013 |

The RMST method yields a bigger p-value than the Cox reg HR. The first occurrence of any composite event can be delayed by 7.8 days on average over 4 years for the patients who are on the Drug C from the patients who are on Warfarin.

# Case IV


Strata arm=0 arm=1
Survival probability
1.00
0.95
0.90
0.85
0.80
0.75
0
100
200
300
400
500
600
700
800
900
Time(Days)
Number at risk
Strata
arm=0 5112 4404 3795 3038 2390 1704 1079 520 112 1
arm=1 10225 8722 7424 5884 4627 3281 2083 1024 222 3
0
100
200
300
400
500
600
700
800
900
Time(Days)

# Case IV


Case IV
Difference in RMST
Estimate and CI
0
5
10
15
0
200
400
600
800
Study Time (Day)
Hazard Ratios
Estimate HR and CI
0.5
1.0
1.5
2.0
0
200
400
600
800
Study Time (Day)
p-values for RMST (Difference) and HR
p values
0.0
0.1
0.2
0.3
0.4
RMST
HR
0
200
400
600
800
Study Time (Day)

# Case IV

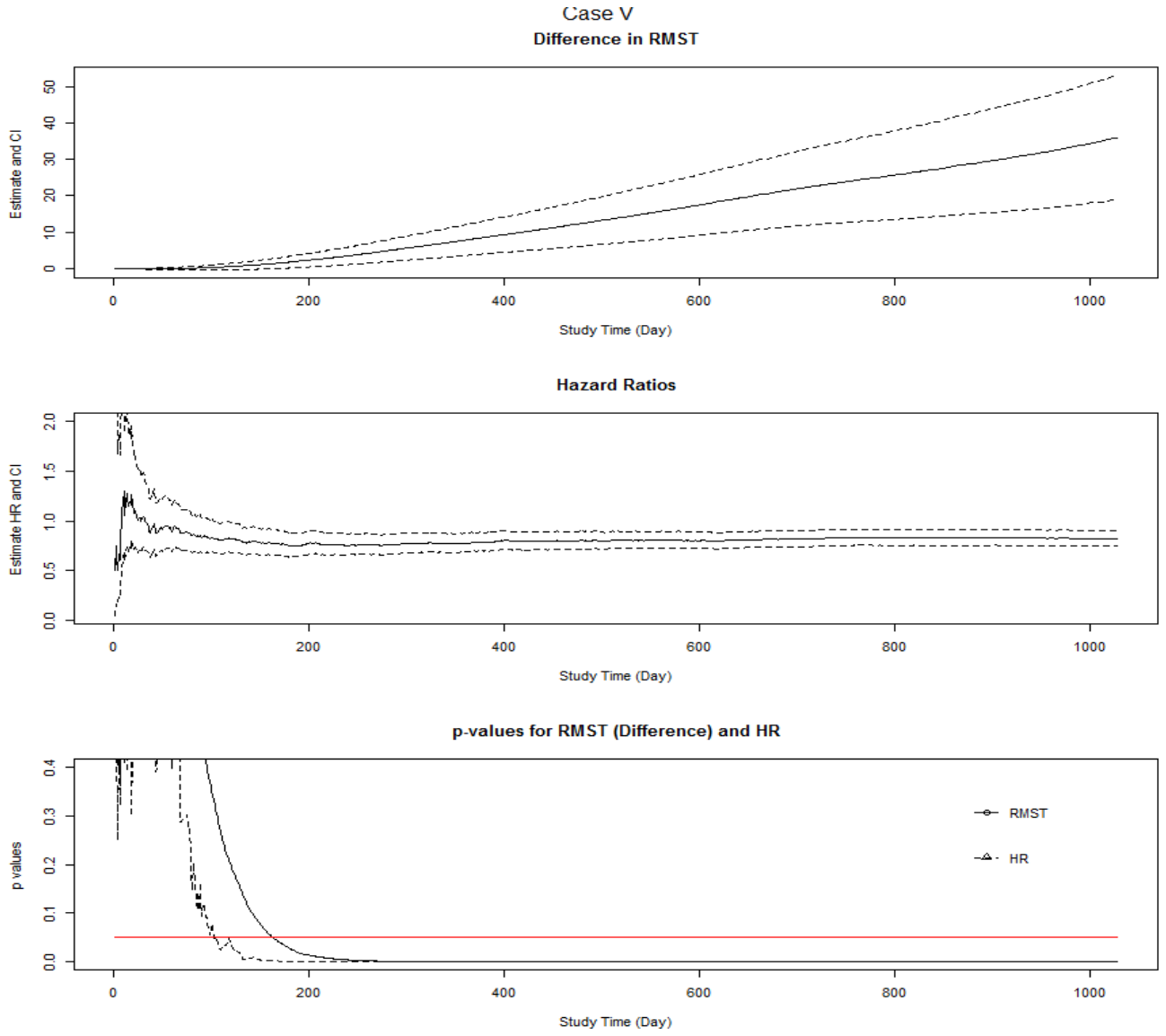
| | Estimate | 95% Lower | 95% Upper | p value |
|---|---|---|---|---|
| Diff in RMST | 10.39 | 2.66 | 18.13 | 0.0084 |
| Cox Reg HR | 0.84 | 0.74 | 0.96 | 0.010 |

The RMST method provides a slightly smaller p-value than the Cox reg HR. The first occurrence of any composite event can be delayed by 10.4 days on average over 3 years for the patients who are on the Drug D.

# Case V


Strata arm=0 arm=1
Survival probability
1.0
0.9
0.8
0.7
0.6
0 100 200 300 400 500 600 700 800 900 1000 1100 1200
Time(Days)
Number at risk
Strata
arm=0 3262 3029 2843 2640 2411 2178 1843 1304 823 462 187 13 4
arm=1 3238 3033 2897 2727 2512 2304 1965 1421 874 481 167 15 3
0 100 200 300 400 500 600 700 800 900 1000 1100 1200
Time(Days)

# Case V

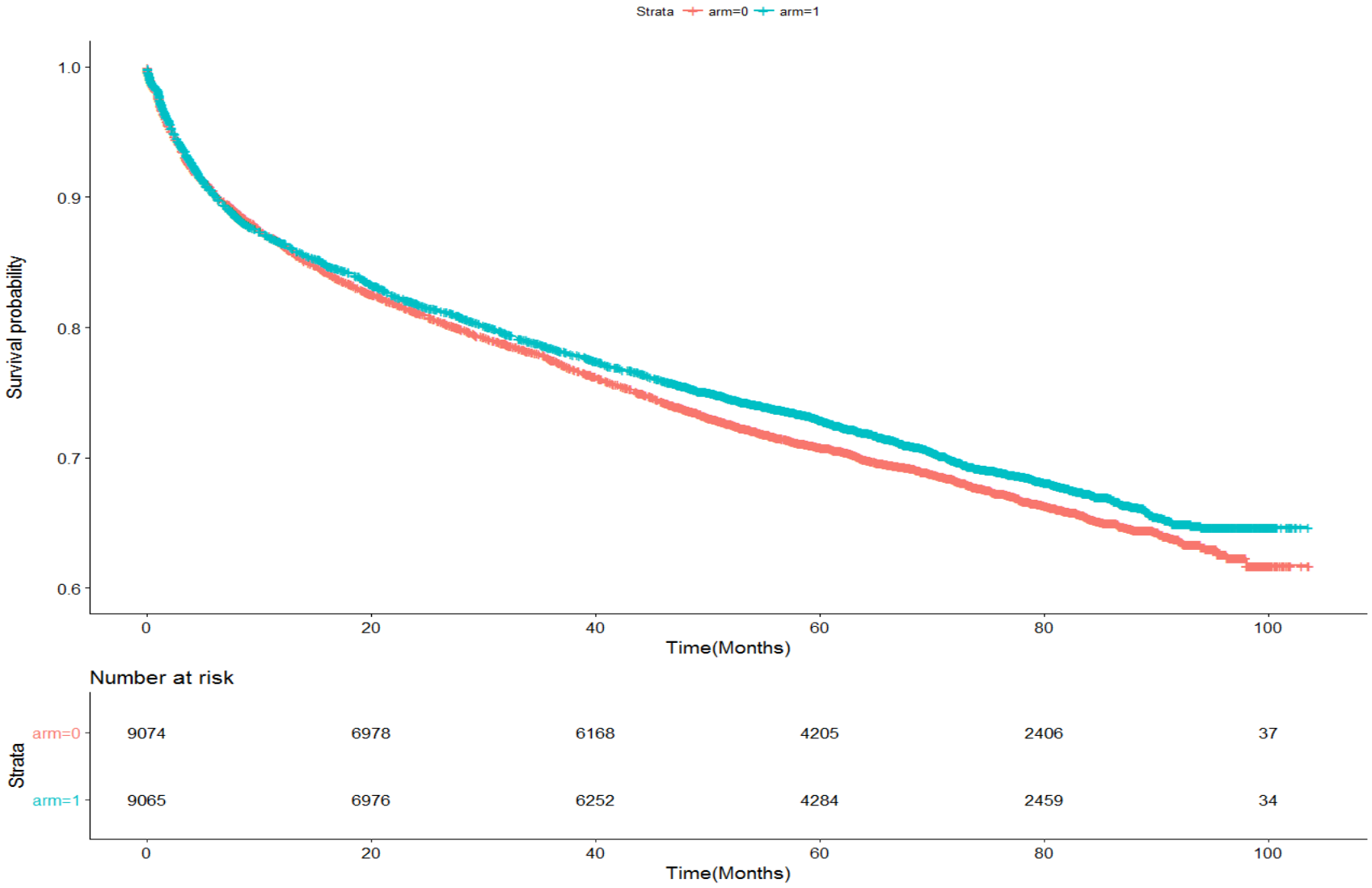

Case V
Difference in RMST
Estimate and CI
0
10
20
30
40
50
0
200
400
600
800
1000
Study Time (Day)
Hazard Ratios
Estimate HR and CI
0.0
0.5
1.0
1.5
2.0
0
200
400
600
800
1000
Study Time (Day)
p-values for RMST (Difference) and HR
p values
0.0
0.1
0.2
0.3
0.4
RMST
HR
0
200
400
600
800
1000
Study Time (Day)

# Case V

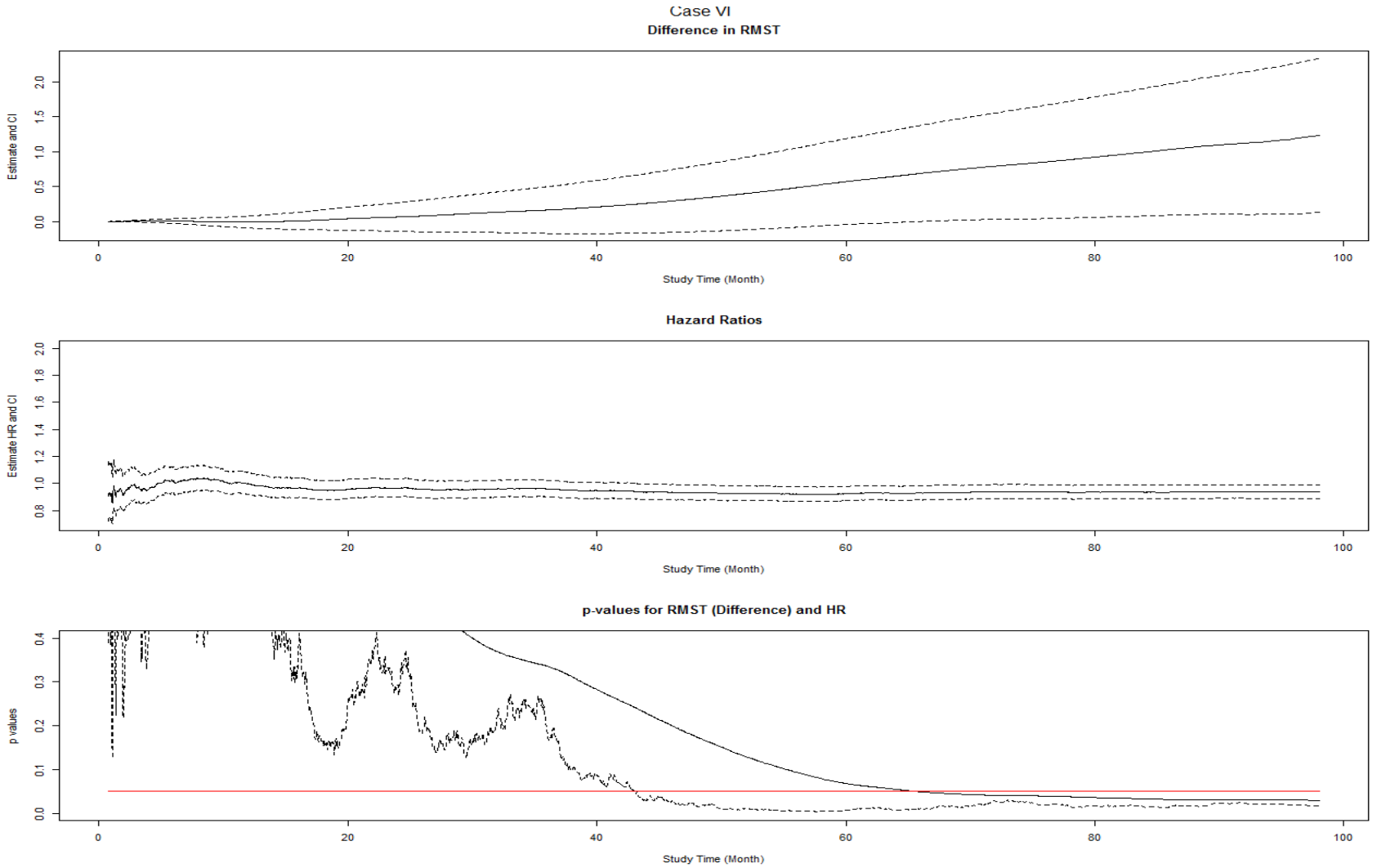
| | Estimate | 95% Lower | 95% Upper | p value |
|---|---|---|---|---|
| Diff in RMST | 35.93 | 18.63 | 52.99 | 0e+00 |
| Cox Reg HR | 0.82 | 0.75 | 0.90 | 1e-04 |

The RMST method provides a smaller p-value than the Cox reg HR. The first occurrence of any composite event can be delayed by 35.7 days on average over 3 years for the patients who are on the Drug E.

# Case VI


Strata arm=0 arm=1
Survival probability
1.0
0.9
0.8
0.7
0.6
0
20
40
60
80
100
Time(Months)
Number at risk
Strata
arm=0
arm=1
9074
6978
6168
4205
2406
37
9065
6976
6252
4284
2459
34
0
20
40
60
80
100
Time(Months)

# Case VI


Case VI
Difference in RMST
Estimate and CI
0.0
0.5
1.0
1.5
2.0
0
20
40
60
80
100
Study Time (Month)
Hazard Ratios
Estimate HR and CI
0.8
1.0
1.2
1.4
1.6
1.8
2.0
0
20
40
60
80
100
Study Time (Month)
p-values for RMST (Difference) and HR
p values
0.0
0.1
0.2
0.3
0.4
0
20
40
60
80
100
Study Time (Month)

# Case VI

| | Estimate | 95% Lower | 95% Upper | p value |
|---|---|---|---|---|
| Diff in RMST | 1.23 | 0.13 | 2.33 | 0.0284 |
| Cox Reg HR | 0.94 | 0.89 | 0.99 | 0.0182 |

The RMST method provides a bigger p value than the Cox reg. The first occurrence of any composite event can be delayed by 1.2 months on average over 8 years for the patients who are on the Drug F & G.

# Remarks

- RMST which is directly related to patient's survival/event-free time, is viable for quantifying treatment effect.
- RMST can give better clinical interpretation of  treatment effect.

# References


- Irwin, J.O., 1949. The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice. *The Journal of hygiene* 47(2), p.188.
- Lawrence, J. 2002. Strategies for changing the test statistic during a clinical trial. *Journal of biopharmaceutical statistics* 12.2: 193-205.
- Schoenfeld, D. 1981. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* 68.1: 316-319.
- Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, Schrag D, Takeuchi M, Uyama Y, Zhao L, Skali H, Solomon S, Jacobus S, Hughes M, Packer M, Wei LJ. 2014. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *Journal of clinical Oncology* 32, 2380-2385.


FDA
U.S. FOOD & DRUG
ADMINISTRATION