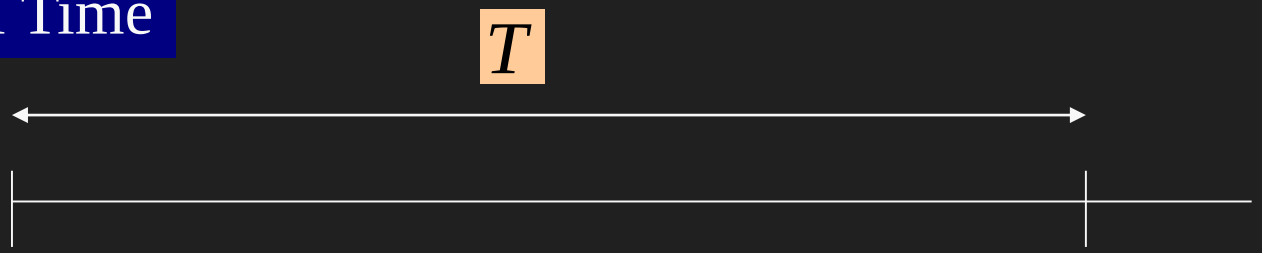
Survival Data Visualization

STAT3622

Survival Analysis

Analysis of time-to-event data

Survival Time



Time origin

End-point

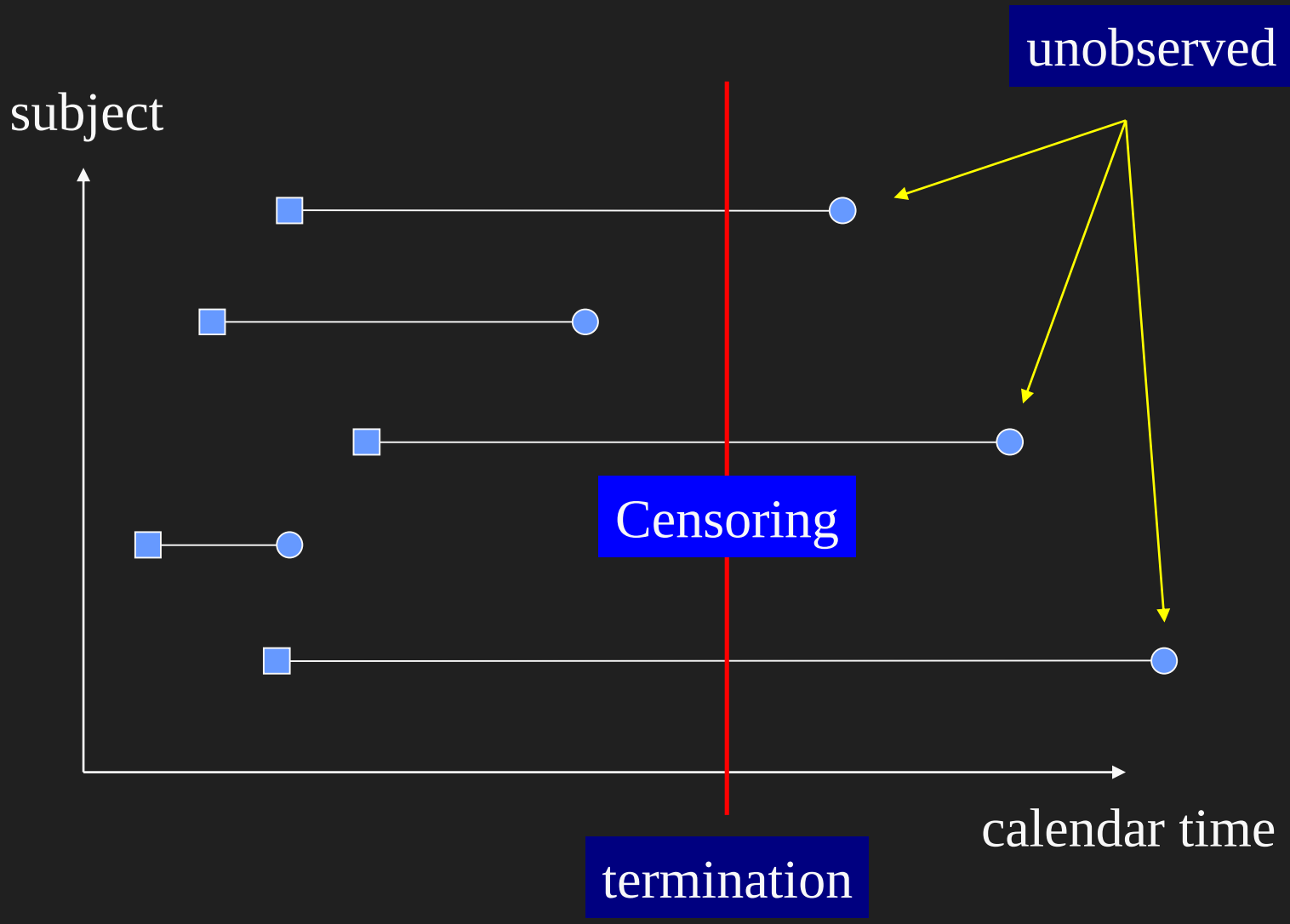
symptom disappear

remission duration

symptom reoccur

Subject-specific

Survival Time



Censoring

End-point unobserved

- termination of study

  funding problem, deadline of reporting

- lost-to-follow-up

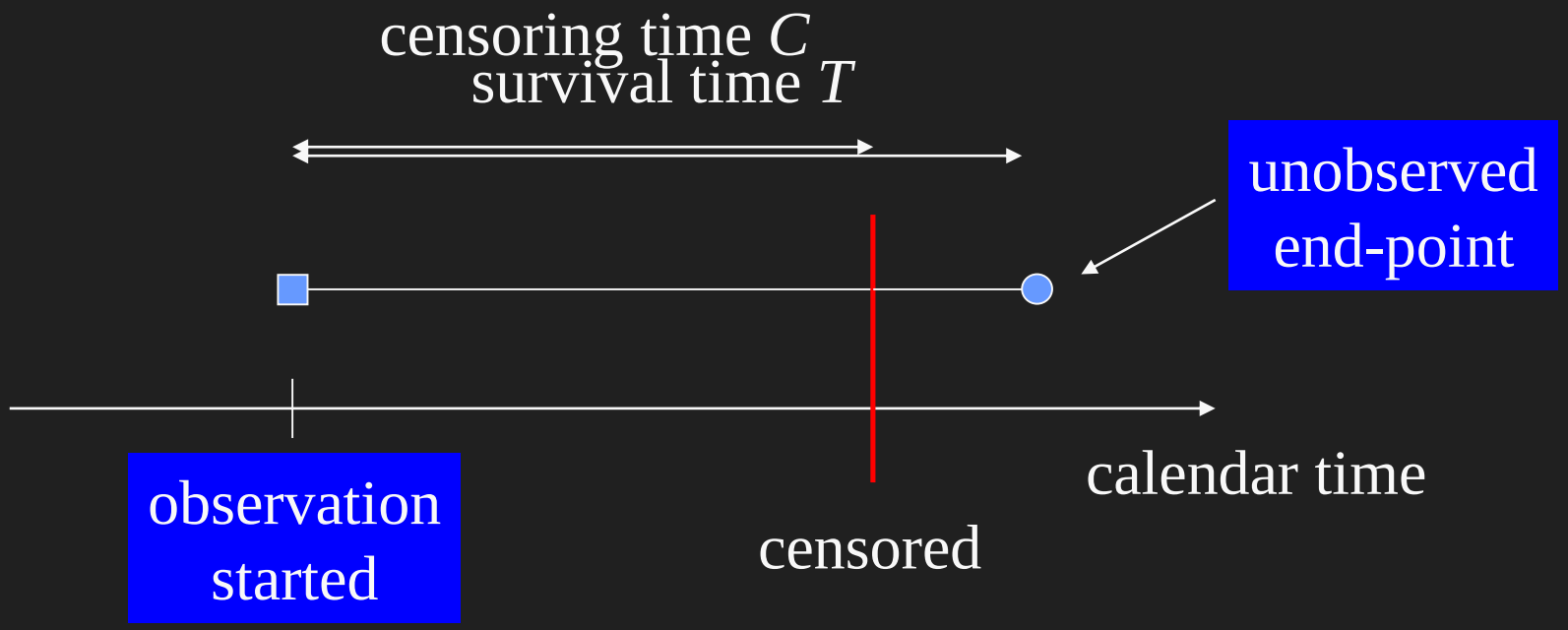
  subject moved or refused to participate

- death due to unrelated cause

  traffic accident on cancer patient

Right Censoring

Censored after the individual entered the study

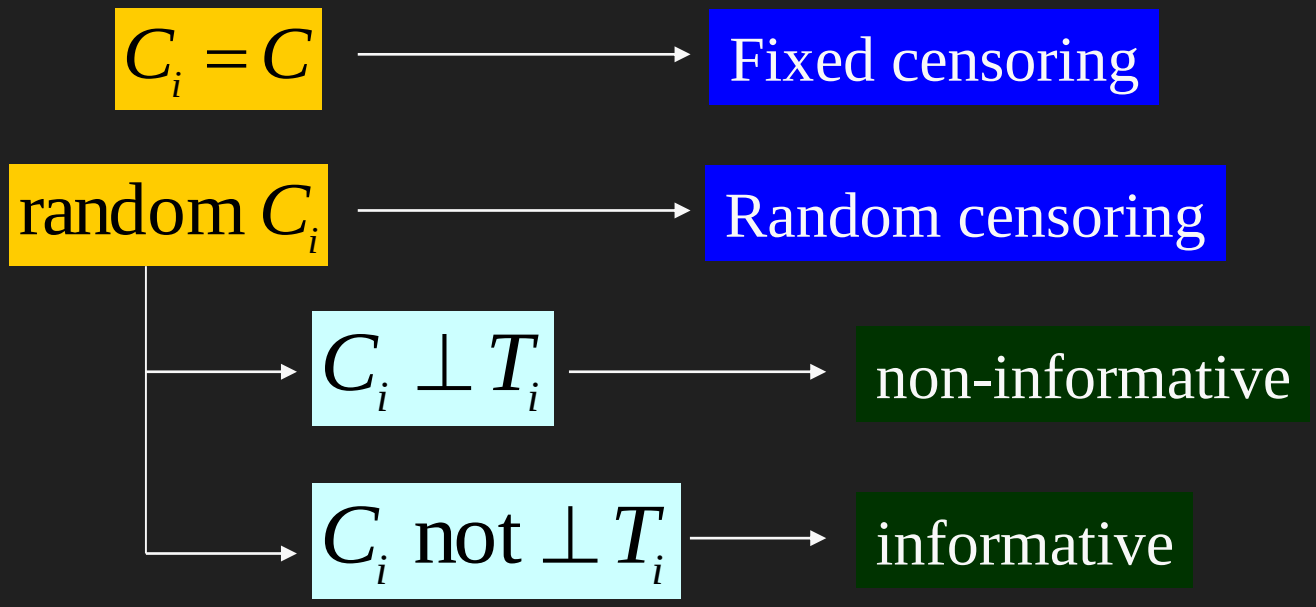


$$C < T$$

Survival Data

T_i —— survival time of i^{th} subject

C_i —— censoring time of i^{th} subject



Survival Data

Complete data $(T_1, C_1), (T_2, C_2), \ldots, (T_n, C_n)$

Right censoring $\longrightarrow$ T_i unobservable if $T_i > C_i$

Observed data $(Y_1, \delta_1), (Y_2, \delta_2), \ldots, (Y_n, \delta_n)$

$$Y_i = \min(T_i, C_i)$$

$$\delta_i = \begin{cases} 1 & \text{if data is uncensored, } T_i \le C_i \\ 0 & \text{if data is censored, } T_i > C_i \end{cases}$$

Survival Function

T —— non-negative random variable with $f(t)$, $F(t)$

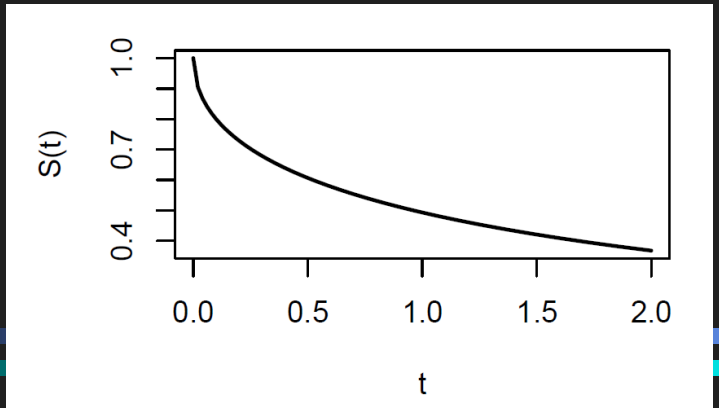
Survivor function (Survival function)

$$S(t) = \Pr(T > t) = 1 - F(t) \quad , t > 0$$

survive at least for time t

Non-increasing

Non-decreasing



Hazard Function

t-units old item $T > t$

t-units old item fail/die at time *t*

$$\Pr(t < T \le t + \Delta t \mid T > t) = \frac{\Pr(t < T \le t + \Delta t)}{\Pr(T > t)}$$

$$= \frac{F(t + \Delta t) - F(t)}{S(t)}$$

Hazard Function

$$\frac{\Pr\left(t < T \le t + \Delta t \mid T > t\right)}{\Delta t} = \frac{1}{S(t)} \frac{F(t+\Delta t) - F(t)}{\Delta t}$$

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr\left(t < T \le t + \Delta t \mid T > t\right)}{\Delta t} = \frac{f(t)}{S(t)}$$

hazard function

Hazard Function

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{1}{S(t)}\frac{dS(t)}{dt} = -\frac{d}{dt}\{\log S(t)\}$$

Cumulative hazard function

$$H(t) = -\log S(t) = \int_0^t \lambda(u)du \quad , t > 0$$

$$S(t) = \exp\{-H(t)\}$$

$$F(t) = 1 - \exp\{-H(t)\}$$

Exponential Distribution

$$F(t) = 1 - \exp(-\lambda t) \quad , t > 0$$

$$H(t) = \lambda t$$

$$\lambda(t) = H'(t) = \lambda$$

constant hazard rate

→ old subjects are likely to die as young subjects

Weibull Distribution

Flexible and simple model

$$\lambda(t) = \alpha\lambda t^{\alpha-1}$$

shape parameter

- $\alpha < 1$ — decreasing
- $\alpha = 1$ — constant
- $\alpha > 1$ — increasing

scale parameter

- larger λ — higher mortality rate

Weibull Distribution

$$\lambda(t) = \alpha\lambda t^{\alpha-1} \qquad H(t) = \int_0^t \alpha\lambda u^{\alpha-1} du \qquad H(t) = \lambda t^{\alpha}$$

$$S(t) = \exp\{-\lambda t^{\alpha}\} \quad , t > 0$$

$$F(t) = 1 - \exp\{-\lambda t^{\alpha}\} \quad , t > 0$$

$$f(t) = \alpha\lambda t^{\alpha-1} \exp\{-\lambda t^{\alpha}\} \quad , t > 0$$

$$T \sim Weibull(\alpha, \lambda)$$

Median Life

100pth percentile

$$x_p = \inf\{x : F(x) \geq p\}$$

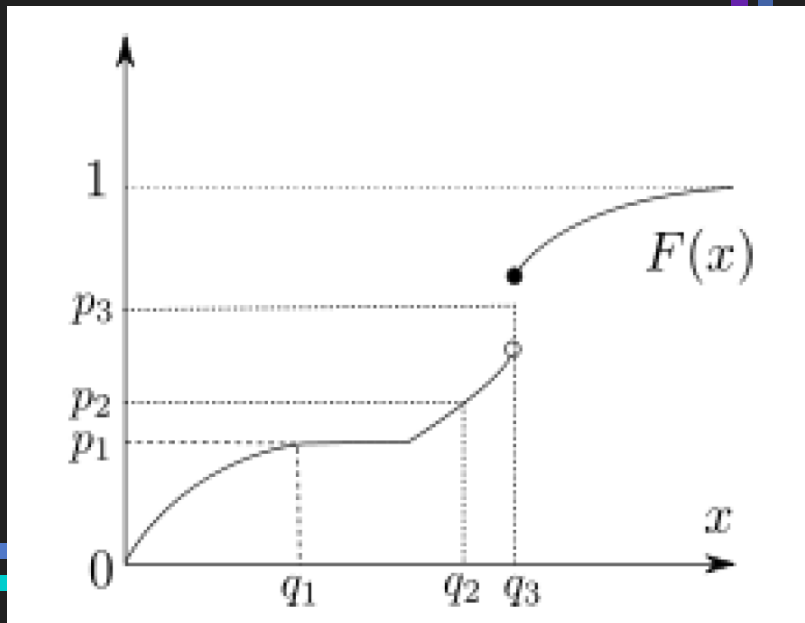
$$x_p = \inf\{x : S(x) \leq 1 - p\}$$

continuous T

$$x_p = S^{-1}(1-p)$$

median lifetime

$$x_{0.5} = S^{-1}(0.5)$$



Kaplan-Meier Estimator

Kaplan-Meier estimator / Product-Limit estimator

$$\hat{S}(t) = 1 \qquad \text{for } t < t_{(1)}$$

$$\hat{S}(t) = \prod_{i=1}^{j} \left(1 - \frac{d_i}{n_i}\right) \qquad \text{for } t_{(j)} \le t < t_{(j+1)}$$

Largest observation is $t_{(k)}$

$$\Rightarrow \hat{S}(t) = 0 \quad \text{for } t \ge t_{(k)}$$

Largest observation is censored at c

$$\Rightarrow \hat{S}(t) = \textit{undefined} \quad \text{for } t \ge c$$

Variance of Kaplan-Meier Estimator

$$\operatorname{var}\left(\hat{\lambda}_j\right) \approx \frac{d_j\left(n_j - d_j\right)}{n_j^3}$$

$$S\left(t_{(j)}\right) = \prod_{i=1}^{j}\left(1-\lambda_i\right)$$

$$\log S\left(t_{(j)}\right) = \sum_{i=1}^{j}\log\left(1-\lambda_i\right)$$

$$\operatorname{var}\left(\log \hat{S}\left(t_j\right)\right) \approx \sum_{i=1}^{j}\frac{1}{\left(1-\hat{\lambda}_i\right)^2} Var\left(\hat{\lambda}_i\right)$$

$$= \sum_{i=1}^{j}\frac{d_i}{n_i\left(n_i - d_i\right)}$$

Standard Error

Greenwood's formula

$$se\left(\hat{S}(t)\right) = \hat{S}(t)\sqrt{\sum_{i=1}^{j} \frac{d_i}{n_i\left(n_i - d_i\right)}}$$

$$t_{(j)} \le t < t_{(j+1)}$$

Complementary Log-Log Transformation

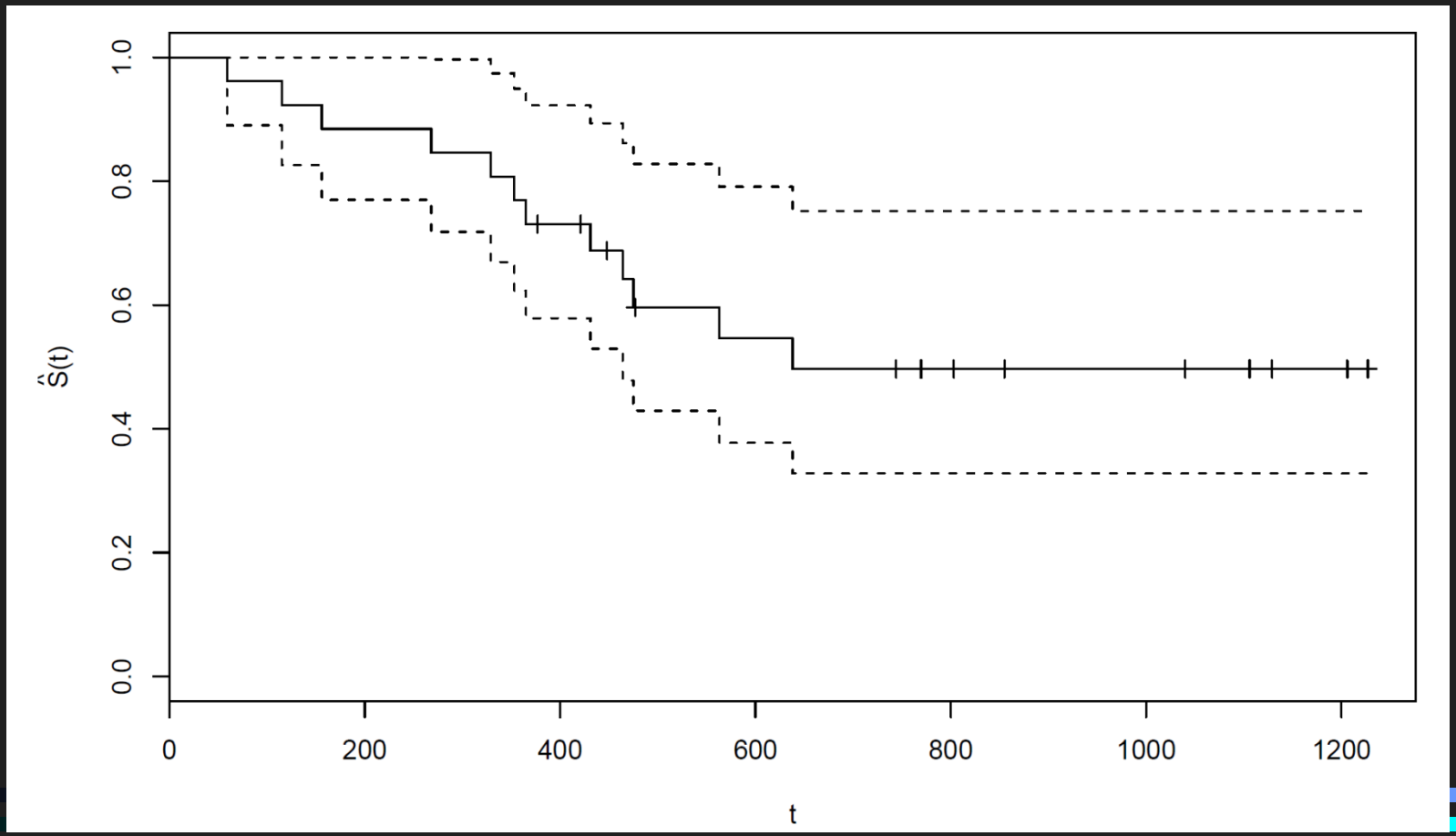
$$\log\left(-\log \hat{S}(t)\right)$$

$$se\left(\log\left(-\log \hat{S}\left(t_{(j)}\right)\right)\right) = \frac{1}{-\log \hat{S}\left(t_{(j)}\right)}\sqrt{\sum_{i=1}^{j} \frac{d_i}{n_i\left(n_i - d_i\right)}}$$

Confidence Interval

100 (1 – α)% C.I. for $S(t)$

$$\hat{S}(t)^{\exp\left(\pm Z_{\alpha/2} se\left(\log\left(-\log \hat{S}(t)\right)\right)\right)}$$



Kaplan-Meier Estimator

Example: 6-MP treated children with acute leukemia

6	6	6	6^+	7	9^+	10	10^+	11^+	13	16
17^+	19^+	20^+	22	23	25^+	32^+	32^+	34^+	35^+	

$t_{(j)}$	d_j	n_j
6	3	21
7	1	17
10	1	15
13	1	12
16	1	11
22	1	7
23	1	6

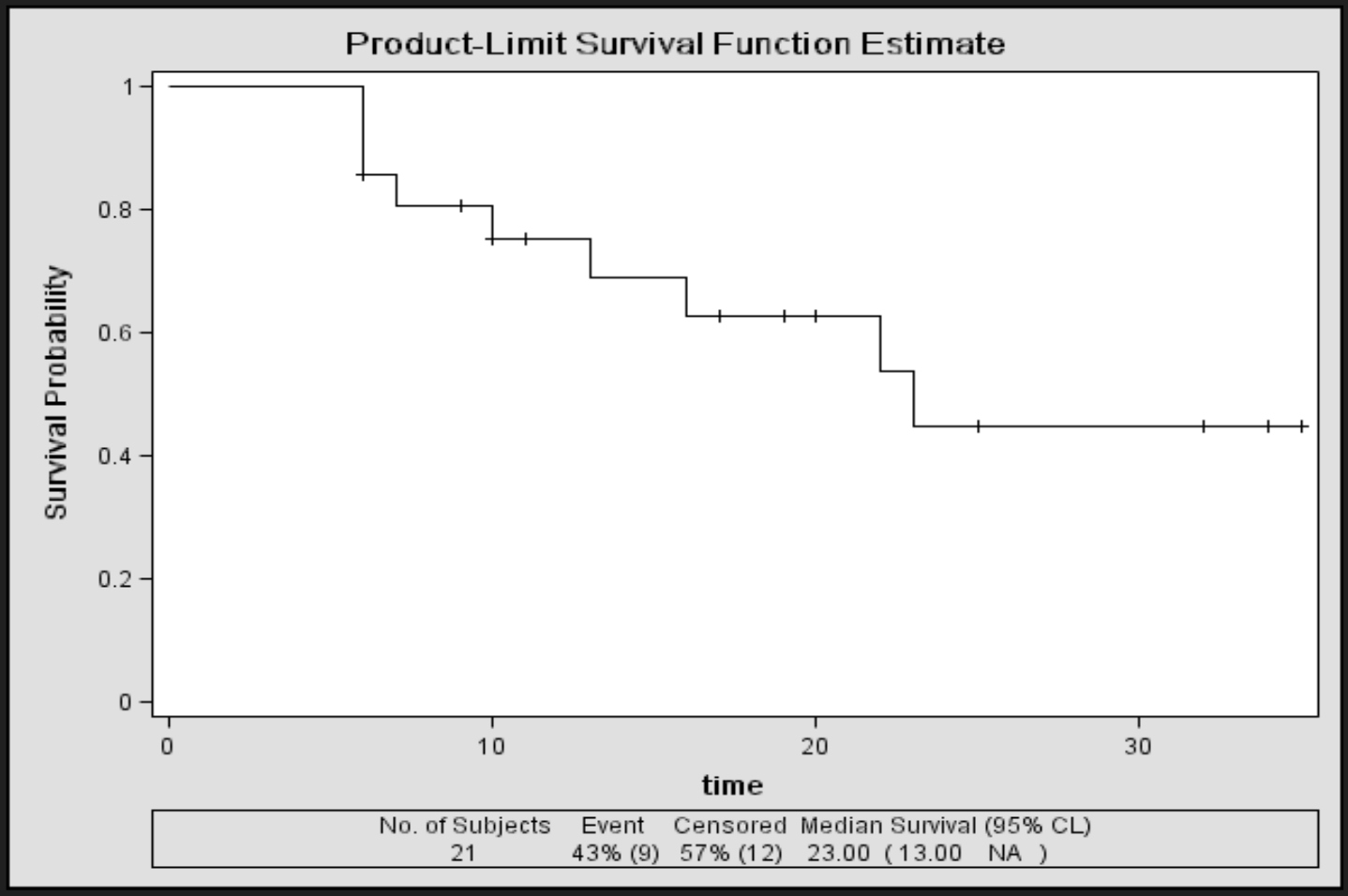
Kaplan-Meier Estimator

$t_{(j)}$	d_j	n_j	$\hat{S}(t)=\prod_{t_{(j)}\le t}\left(1-\frac{d_j}{n_j}\right)$	$\sum_{t_{(j)}\le t}\frac{d_j}{n_j(n_j-d_j)}$	$\hat{S}(t)^2\sum_{t_{(j)}\le t}\frac{d_j}{n_j(n_j-d_j)}$
6	3	21	$\hat{S}(6)=0.857$	0.0079	0.0058
7	1	17	0.807	0.0116	0.0076
10	1	15	0.753	0.0164	0.0093
13	1	12	0.690	0.0240	0.0114
16	1	11	0.628	0.0330	0.0130
22	1	7	0.538	0.0569	0.0164
23	1	6	0.448	0.0902	0.0181

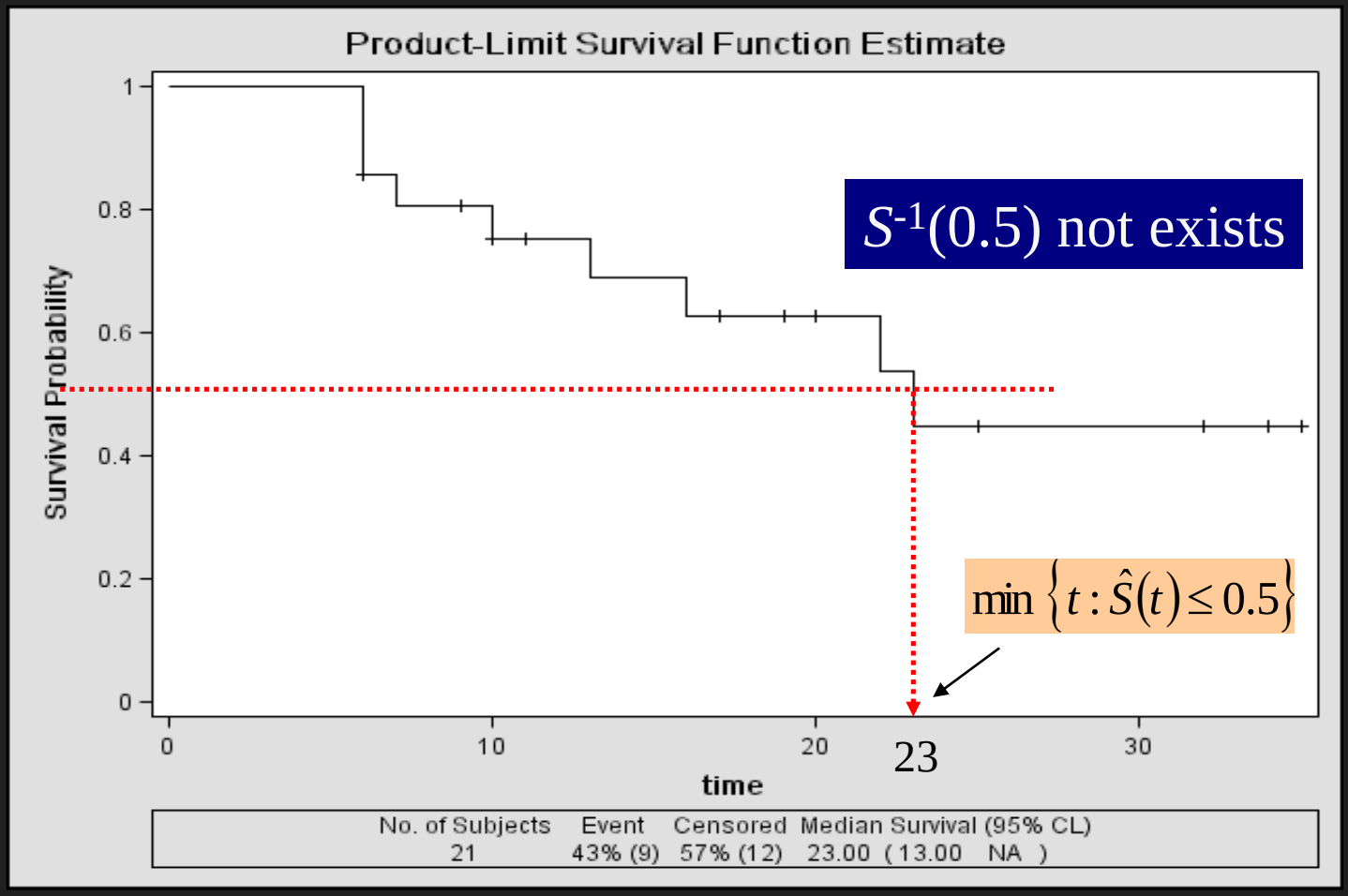
C.I. for S(6)

$$(0.857)^{\exp\left(\pm 1.96\times\frac{\sqrt{0.0079}}{-\log 0.857}\right)} = [0.621, 0.951]$$

Kaplan-Meier Estimator



Estimation of Percentiles



Nelson-Aalen Estimator

$$S(t) = \exp(-H(t))$$

$$\hat{H}(t) = -\sum_{i=1}^{j} \log\left(1 - \frac{d_i}{n_i}\right) \approx \sum_{i=1}^{j} \frac{d_i}{n_i}$$

$$t_{(j)} \le t < t_{(j+1)}$$

Nelson-Aalen estimate of cumulative hazard function

$$\hat{H}(t) = \sum_{i=1}^{j} \frac{d_i}{n_i} \qquad \text{for } t_{(j)} \le t \le t_{(j+1)}$$

Nelson-Aalen Estimator

Nelson-Aalen estimate of cumulative hazard function

$$\hat{H}(t) = \sum_{i=1}^{j} \frac{d_i}{n_i} \qquad \text{for } t_{(j)} \le t \le t_{(j+1)}$$

$$se\left(\hat{H}(t)\right) = \sqrt{\sum_{i=1}^{j} \frac{d_i}{n_i^2}}$$

Nelson-Aalen estimate of survivor function

$$\hat{S}(t) = \exp\left(-\hat{H}(t)\right)$$

$$se\left(\hat{S}(t)\right) = \hat{S}(t)\sqrt{\sum_{i=1}^{j} \frac{d_i}{n_i^2}}$$

Nelson-Aalen Estimator

Complementary Log-Log Transformation

$$\log\left(-\log \hat{S}(t)\right)$$

$$se\left(\log\left(-\log \hat{S}\left(t_{(j)}\right)\right)\right) = \frac{1}{-\log \hat{S}\left(t_{(j)}\right)} \sqrt{\sum_{i=1}^{j} \frac{d_i}{n_i^2}}$$

100 (1 – α)% C.I. for $S(t)$

$$\hat{S}(t)^{\exp\left(\pm Z_{\alpha/2} se\left(\log\left(-\log \hat{S}(t)\right)\right)\right)}$$

Nelson-Aalen Estimator

$t_{(j)}$	d_j	n_j	$\hat{H}(t)=\sum_{t_{(j)}\le t}\frac{d_j}{n_j}$	$\mathrm{var}\left(\hat{H}(t)\right)=\sum_{t_{(j)}\le t}\frac{d_j}{n_j^2}$
6	3	21	0.1429	0.0068
7	1	17	0.2017	0.0103
10	1	15	0.2684	0.0147
13	1	12	0.3517	0.0216
16	1	11	0.4426	0.0299
22	1	7	0.5855	0.0503
23	1	6	0.7522	0.0781

$$\frac{3}{}$$

$$201\ 0.0068+\frac{1}{17^2}$$

2684

Nelson-Aalen Estimator

t	$\hat{H}(t)$	$se\left(\hat{H}(t)\right)$	$\hat{S}(t)$	$se\left(\hat{S}(t)\right)$
$0 \le t < 6$	0	0	1	0
$6 \le t < 7$	0.1429	0.0825	0.8668	0.0715
$7 \le t < 10$	0.2017	0.1015	0.8173	0.0830
$10 \le t < 13$	0.2684	0.1212	0.7646	0.0927
$13 \le t < 16$	0.3517	0.1470	0.7035	0.1034
$16 \le t < 22$	0.4426	0.1729	0.6424	0.1111
$22 \le t < 23$	0.5855	0.2243	0.5568	0.1249
$23 \le t < 35$	0.7522	0.2795	0.4713	0.1317
$t \ge 35$	////////////////////	////////////////////	////////////////////	////////////////////

Parametric Model Checking

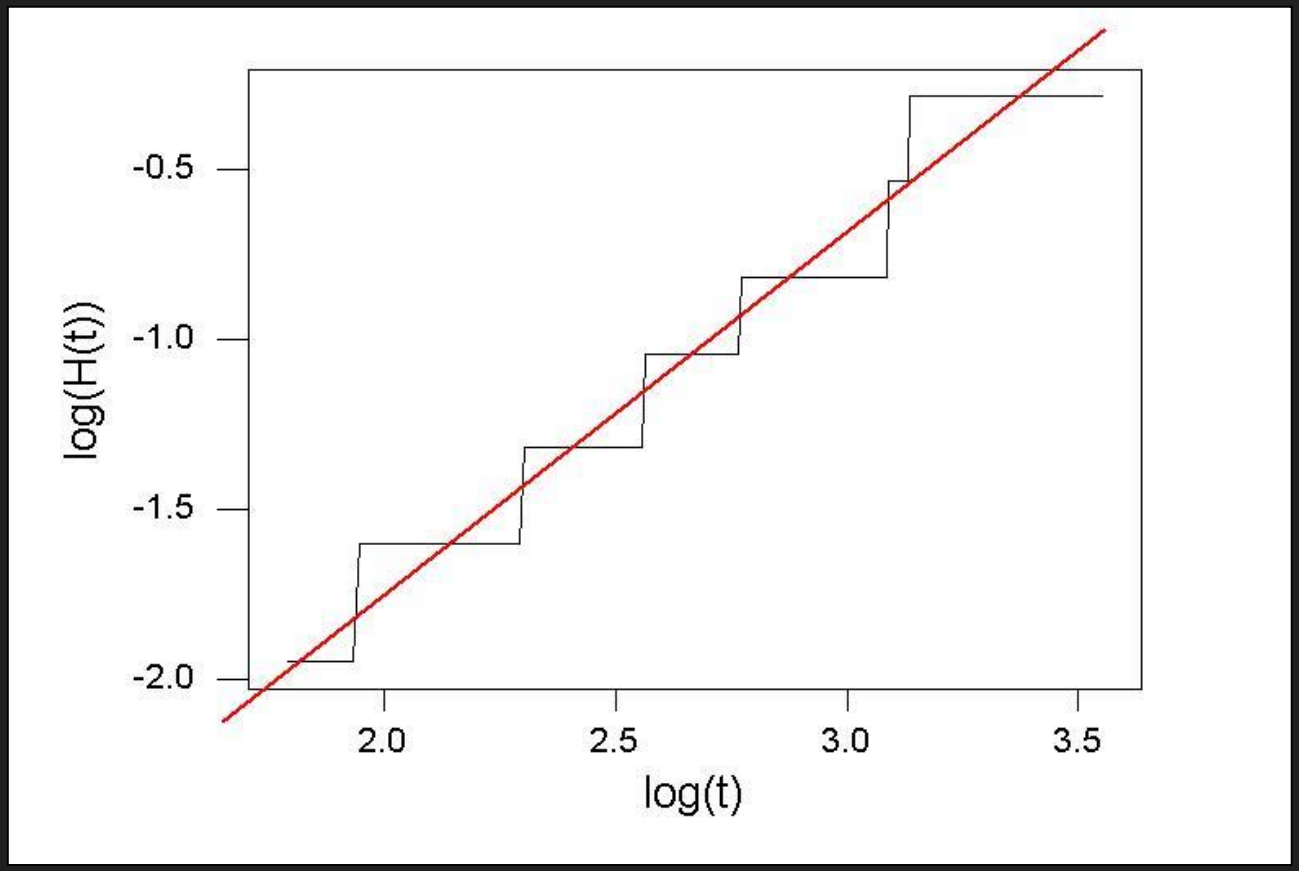
Exponential $\longrightarrow$ $H(t) = \lambda t$

Plot of $\hat{H}(t)$ vs t shows a straight line

Weibull $\longrightarrow$ $\log H(t) = \log \lambda + \alpha \log t$

Plot of $\log \hat{H}(t)$ vs $\log t$ shows a straight line

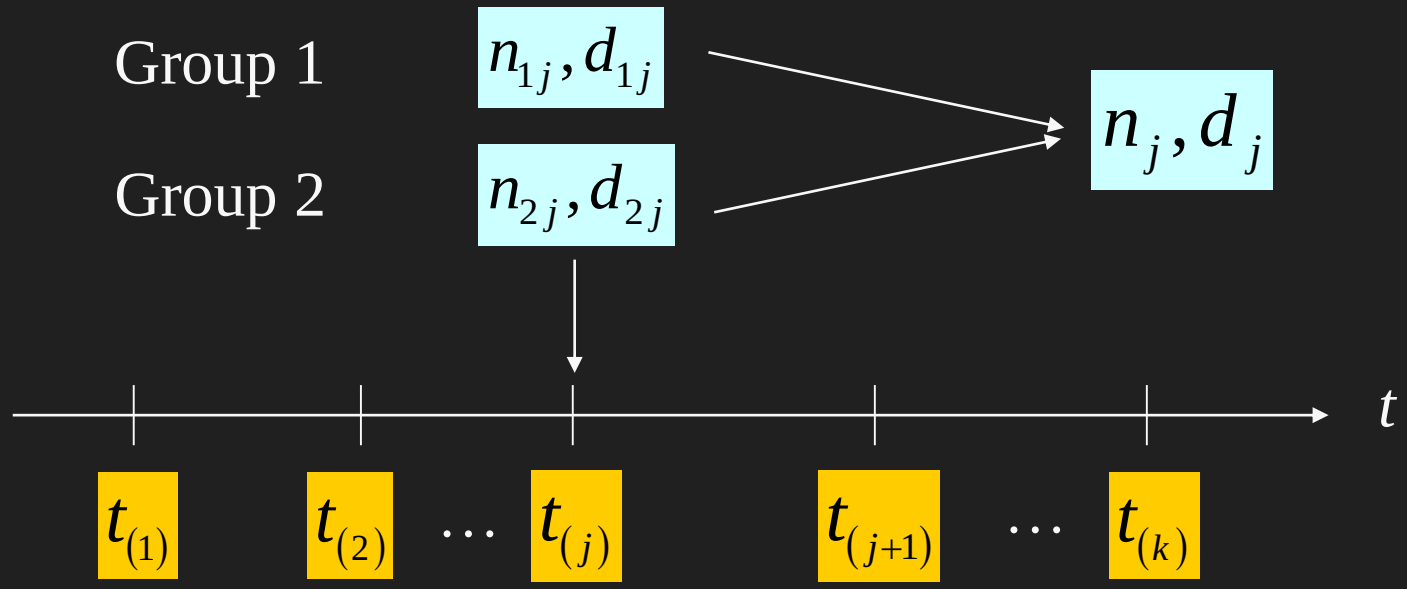
Parametric Model Checking



Two-Samples Comparison

$H_0 : S_1(t) = S_2(t)$ for all $t \geq 0$

$H_1 : S_1(t) \neq S_2(t)$ for some $t \geq 0$



Distinct uncensored survival times of ***pooled sample***

Two-Samples Comparison

Group	Deaths at $t_{(j)}$	Alive beyond $t_{(j)}$	At risk just before $t_{(j)}$
1	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
2	d_{2j}	$n_{2j} - d_{2j}$	n_{2j}
Total	d_j	$n_j - d_j$	n_j

Groups indifference $\longrightarrow$ Group 1 is sample n_{1j} from n_j

$$\hat{e}_{1j} = E\left(d_{1j}\right) = \frac{n_{1j} d_j}{n_j} \qquad \hat{v}_{1j} = Var\left(d_{1j}\right) = \frac{n_{1j} n_{2j} d_j \left(n_j - d_j\right)}{n_j^2 \left(n_j - 1\right)}$$

d_j, n_{1j}

Two-Samples Comparison

$$\hat{e}_{1j} = E\left(d_{1j}\right) = \frac{n_{1j} d_j}{n_j}$$

$$\hat{v}_{1j} = Var\left(d_{1j}\right) = \frac{n_{1j} n_{2j} d_j \left(n_j - d_j\right)}{n_j^2 \left(n_j - 1\right)}$$

Measure of deviation from null hypothesis

$$Z_1 = \sum_{j=1}^{k} \left(d_{1j} - \frac{n_{1j} d_j}{n_j} \right)$$

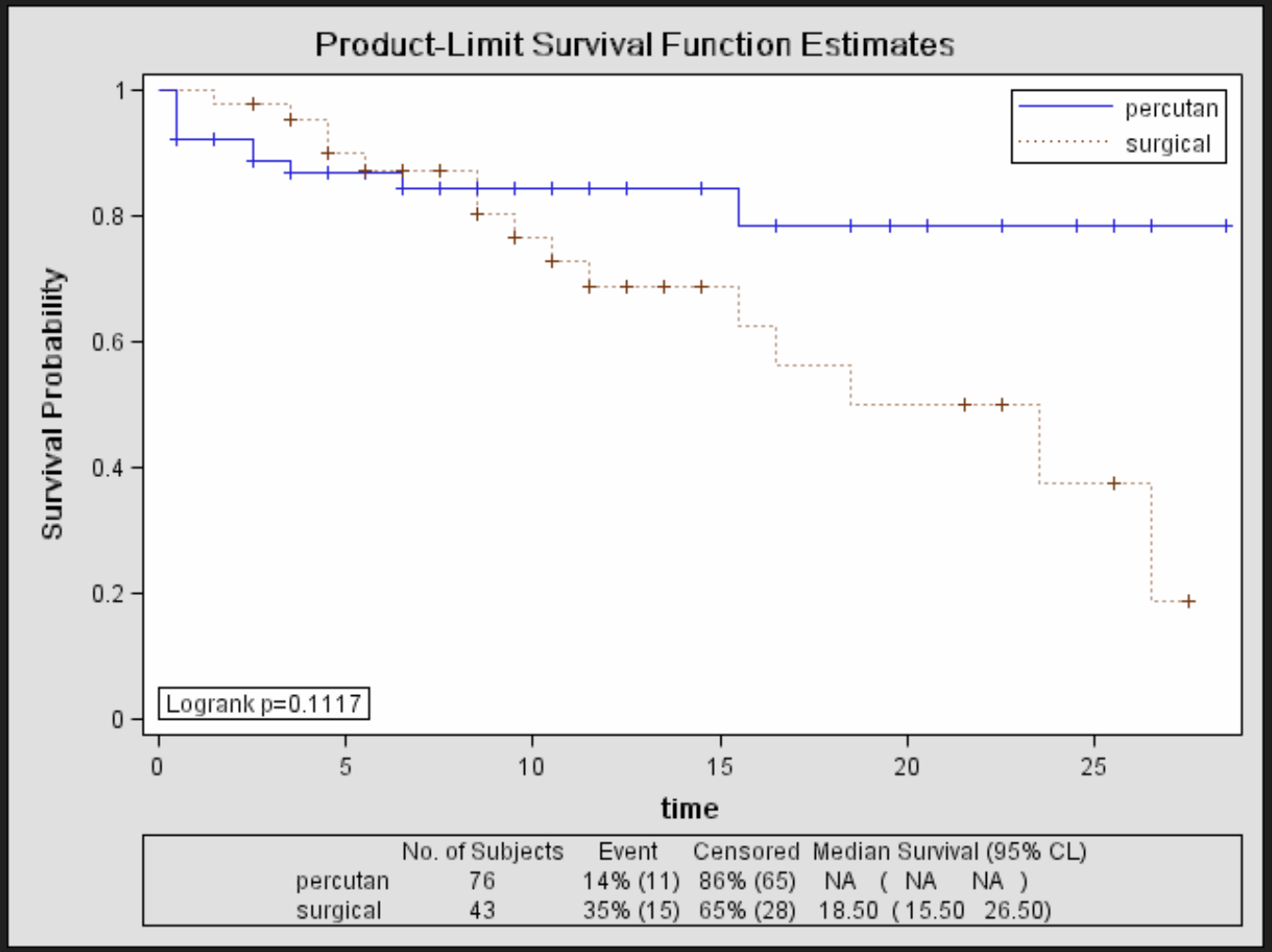
$$Var\left(Z_1\right) = \sum_{j=1}^{k} \hat{v}_j = \sum_{j=1}^{k} \frac{n_{1j} n_{2j} d_j \left(n_j - d_j\right)}{n_j^2 \left(n_j - 1\right)}$$

$$Q = \frac{Z_1^2}{Var\left(Z_1\right)} \overset{H_0}{\sim} \chi_1^2$$

Two-Samples Comparison

Catheter Placed	Time (in months) to cutaneous exit-site infection									
Surgical	1.5	2.5+	2.5+	3.5	3.5+	3.5+	3.5+	4.5	4.5	4.5+
	5.5	5.5+	6.5+	6.5+	7.5+	7.5+	7.5+	7.5+	8.5	8.5
	8.5+	9.5	9.5+	10.5	10.5+	11.5	11.5+	12.5+	12.5+	13.5+
	14.5+	14.5+	15.5	16.5	18.5	21.5+	21.5+	22.5+	22.5+	23.5
	25.5+	26.5	27.5+							
Percutaneous	0.5	0.5	0.5	0.5	0.5	0.5	0.5+	0.5+	0.5+	0.5+
	0.5+	0.5+	0.5+	0.5+	0.5+	0.5+	1.5+	1.5+	1.5+	1.5+
	2.5	2.5	2.5+	2.5+	2.5+	2.5+	2.5+	3.5	3.5+	3.5+
	3.5+	3.5+	3.5+	4.5+	4.5+	4.5+	5.5+	5.5+	5.5+	5.5+
	5.5+	6.5	6.5+	7.5+	7.5+	7.5+	8.5+	8.5+	8.5+	9.5+
	9.5+	10.5+	10.5+	10.5+	11.5+	11.5+	12.5+	12.5+	12.5+	12.5+
	14.5+	14.5+	15.5	16.5+	16.5+	18.5+	19.5+	19.5+	19.5+	20.5+
	22.5+	24.5+	25.5+	26.5+	26.5+	28.5+				

Two-Samples Comparison



Two-Samples Comparison

$t_{(j)}$	n_{1j}	d_{1j}	n_{2j}	d_{2j}	n_j	d_j	$\frac{n_{1j}d_j}{n_j}$	$d_{1j}-\frac{n_{1j}d_j}{n_j}$	$\frac{n_{1j}n_{2j}}{n_j^2}\left(\frac{n_j-d_j}{n_j-1}\right)d_j$
0.5	43	0	76	6	119	6	2.168	-2.168	1.326
1.5	43	1	60	0	103	1	0.417	0.583	0.243
2.5	42	0	56	2	98	2	0.857	-0.857	0.485
…	…	…	…	…	…	…	…	…	…
23.5	4	1	5	0	9	1	0.444	0.556	0.247
26.5	2	1	3	0	5	1	0.400	0.600	0.240
Sum								3.964	6.211

$$Q = \frac{Z_1^2}{Var(Z_1)} = \frac{3.964^2}{6.211} = 2.530 \quad < \quad \chi^2_{1,0.05} = 3.841$$

No significant difference

Two-Samples Comparison

Log-rank test

$$Z_1 = \sum_{j=1}^{k} \left(d_{1j} - \frac{n_{1j} d_j}{n_j} \right)$$

$$Var(Z_1) = \sum_{j=1}^{k} \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$$

General test

$$Z_1 = \sum_{j=1}^{k} w(t_{(j)}) \left(d_{1j} - \frac{n_{1j} d_j}{n_j} \right)$$

$$Var(Z_1) = \sum_{j=1}^{k} \left[w(t_{(j)}) \right]^2 \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$$

Two-Samples Comparison

Test	Weight function $w(t_{(j)})$
Log-Rank	1
Wilcoxon (Gehan)	n_j
Tarone-Ware	$\sqrt{n_j}$
Peto-Peto	$\tilde{S}(t_{(j)}) = \prod_{t_{(i)} \le t_{(j)}} \left(1 - \frac{d_i}{n_i + 1}\right)$
Modified Peto-Peto	$\tilde{S}(t_{(j)}) = \frac{n_j}{n_j + 1}$
Fleming-Harrington	$\hat{S}(t_{(j)})^p \left[1 - \hat{S}(t_{(j)})\right]^q \quad \text{for } p, q \ge 0$

Multiple Comparison

$$H_0 : S_1(t) = S_2(t) = \cdots = S_G(t) \quad \text{for all } t \geq 0$$

$$Z_i = \sum_{j=1}^{k} \left(d_{ij} - \frac{n_{ij} d_j}{n_j} \right) \quad \text{for } i = 1,2,...,G$$

$$Q = \mathbf{Z}' \Sigma \mathbf{Z} \overset{H_0}{\sim} \chi^2_{G-1}$$

Cox Regression Model

Proportional hazards model

$$\lambda_1(t) = c\lambda_0(t)$$

$$\lambda_1(t) = e^{\beta}\lambda_0(t)$$

$$\lambda(t; X) = e^{\beta X}\lambda_0(t)$$

explanatory variable

baseline hazard function

Treatment ($X = 1$)
Control ($X = 0$)

Cox Regression Model

General Proportional hazards model

$$\lambda\left(t; X_1,..., X_p\right) = \exp\left(\beta_1 X_1 + \cdots + \beta_p X_p\right)\lambda_0(t)$$

$$\log \frac{\lambda\left(t; \mathbf{X}\right)}{\lambda\left(t; \mathbf{X}^*\right)} = \beta_1\left(X_1 - X_1^*\right) + \cdots + \beta_p\left(X_p - X_p^*\right)$$

does not depend on *t*

Binary predictor

$$HR_{lung\,cancer\,/\,smoking} = \frac{h_i(t)}{h_j(t)} = \frac{\lambda_0(t)e^{\beta_{smoking}(1)+\beta_{age}(60)}}{\lambda_0(t)e^{\beta_{smoking}(0)+\beta_{age}(60)}} = e^{\beta_{smoking}(1-0)}$$

$$HR_{lung\,cancer\,/\,smoking} = e^{\beta_{smoking}}$$

This is the hazard ratio for smoking adjusted for age.

Continuous predictor

$$HR_{lung\ cancer\ /10-years\ \text{increase in age}} = \frac{h_i(t)}{h_j(t)} = \frac{\lambda_0(t)e^{\beta_{smoking}(0)+\beta_{age}(70)}}{\lambda_0(t)e^{\beta_{smoking}(0)+\beta_{age}(60)}} = e^{\beta_{age}(70-60)}$$

$$HR_{lung\ cancer\ /10-years\ \text{increase in age}} = e^{\beta_{age}(10)}$$

This is the hazard ratio for a 10-year increase in age, adjusted for smoking.

Exponentiating a continuous predictor gives you the hazard ratio for a 1-unit increase in the predictor.

Survival function and hazard function…

$$\text{Survival from hazard}: S(t) = e^{(-\int_0^t h(u)du)}$$

$$h_i(t) = \lambda_0(t)e^{\beta x_i}$$

$$P_i(T > t \mid x) = S_i(t) = e^{(-\int_0^t \lambda_0(u)e^{\beta x}du)}$$

RMST vs HR

- Issues and concerns of HR estimate
- Alternatives to HR
  -- model-free summary measures
- Examples from cancer trials
- Implementation with R
- Conclusions

Issues and concerns about hazard ratio estimate (1)

No reference number

A ratio of 0.8 is difficult to interpret clinically without any absolute hazard to serve as reference

… even if the PH assumption is correct

Issues and concerns about hazard ratio estimate (2)

Decision-making in rare event cases

When the number of events is small, the hazard ratio estimate is very unstable and the confidence interval is wide, implying that there is not enough information to make a decision

… even if the PH assumption is correct

Issues and concerns about hazard ratio estimate (3)

… if the PH assumption is violated

- It is difficult to interpret the parameter being estimated. It is not a simple average of the hazard ratio over time
- The parameter to be estimated depends on underlying study-specific censoring distributions

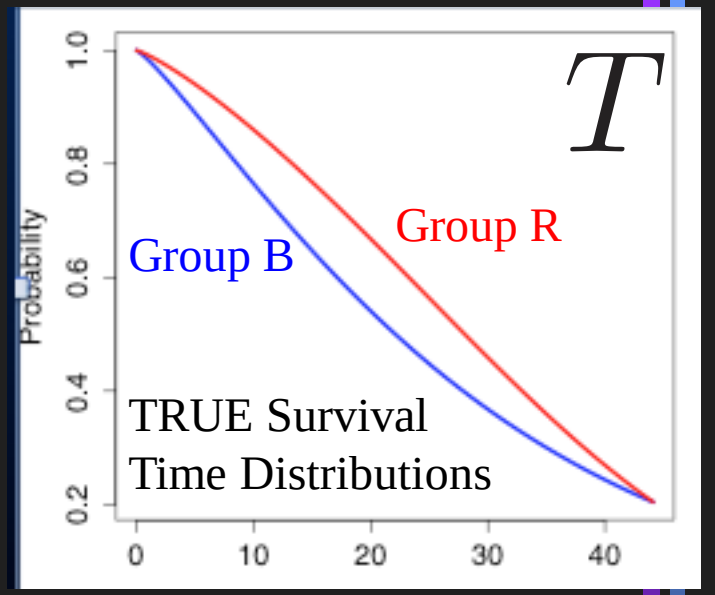
!! Any estimate for a model-based between-group difference metric has the similar issue…

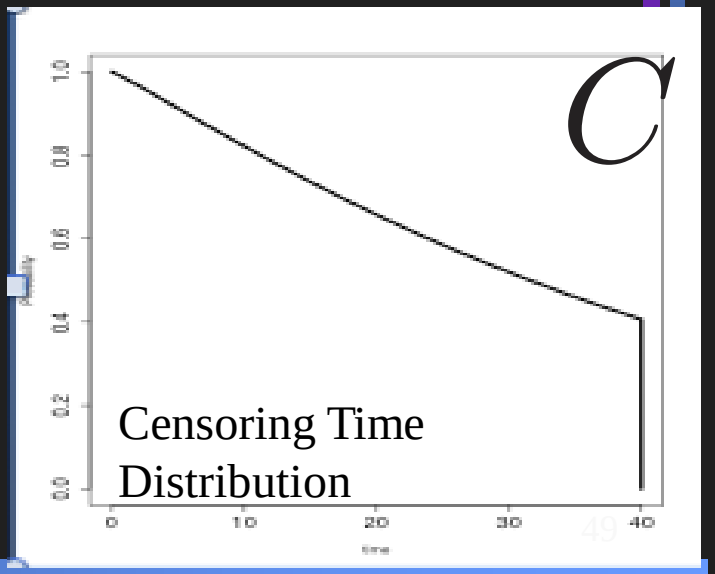
A numerical study for illustration

- Consider two groups and their true survival functions
- Consider a common censoring time distribution
- Generate 1 million of (T, C) for each group and get "observable" survival data (X, Δ)

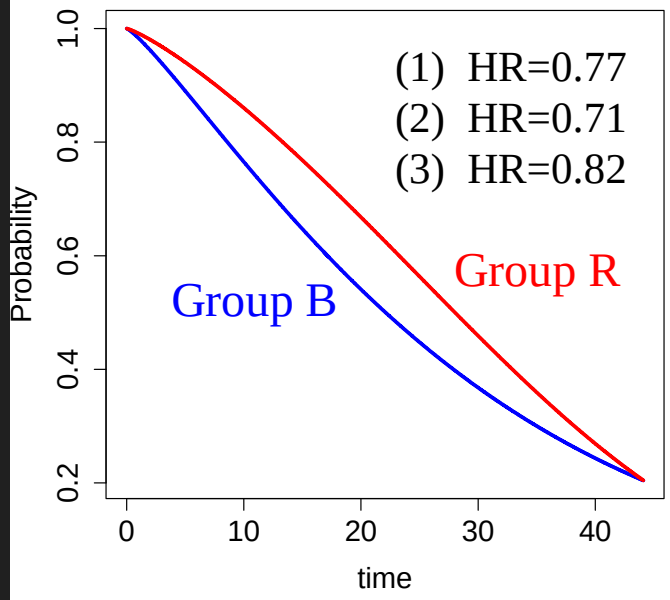
$$X = \min(T, C)$$
$$\Delta = 1 \text{ if } T \leq C$$
$$0 \text{ otherwise}$$

- Calculate the HR by the PH estimator

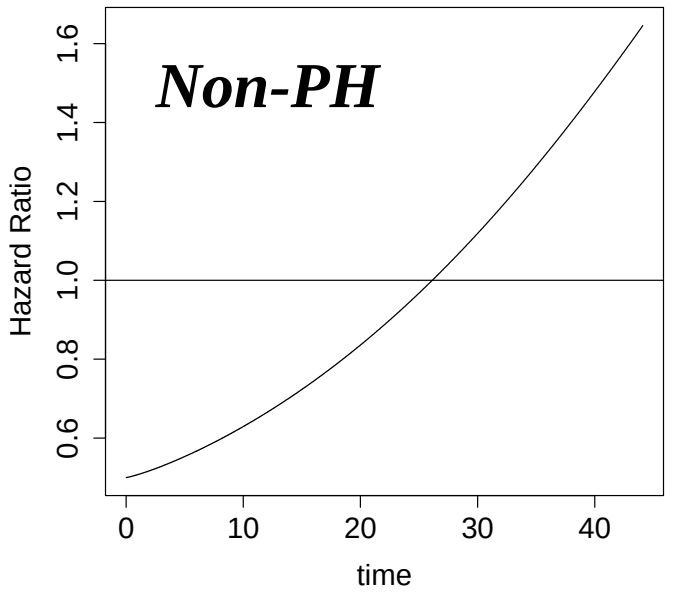




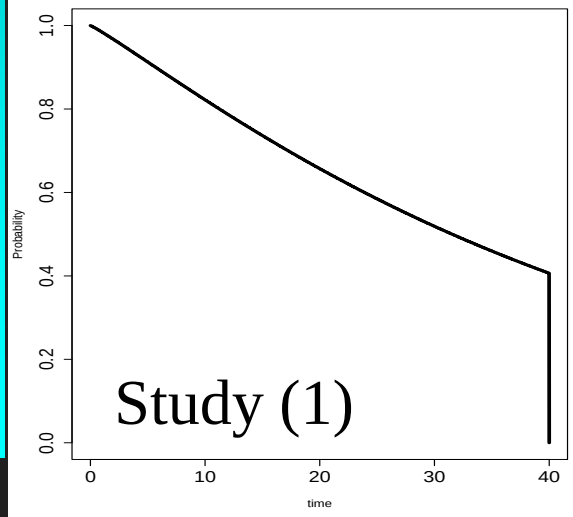
“TRUE” Survival functions

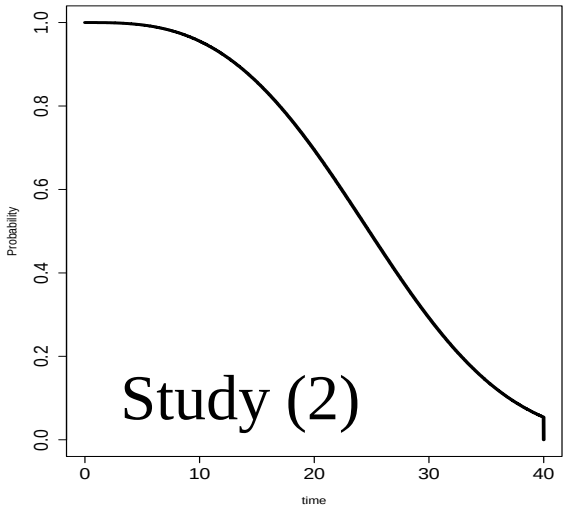


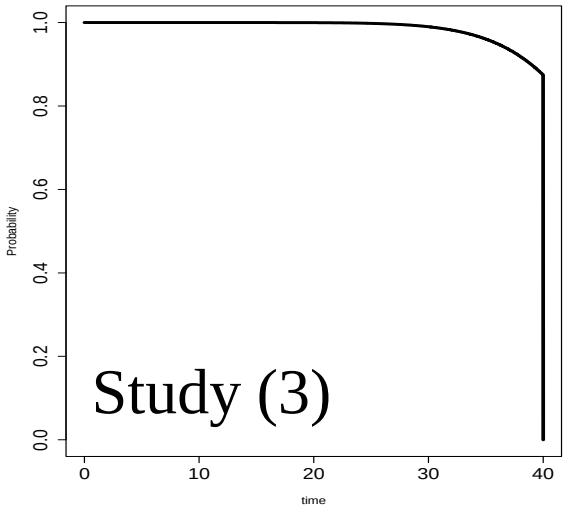
“TRUE” Hazard Ratio



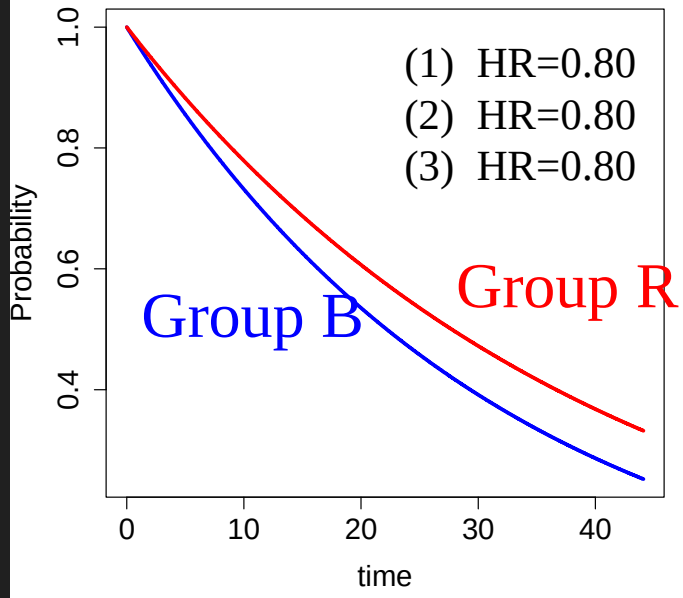
Study-specific censoring distribution



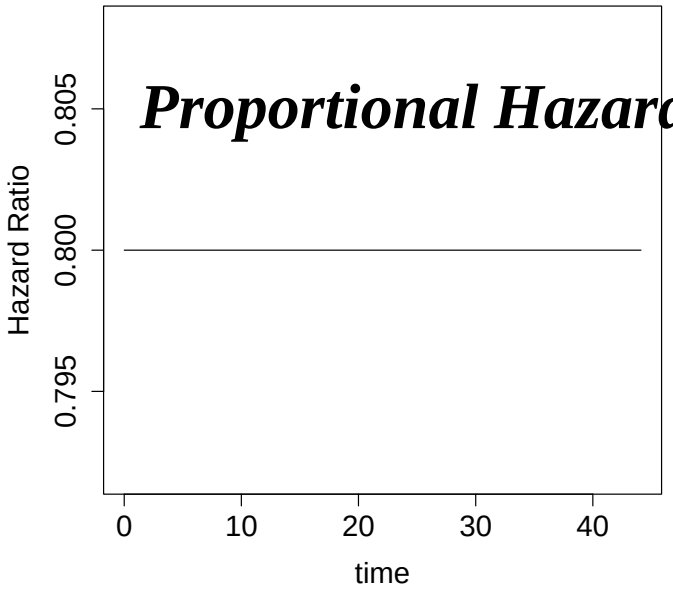




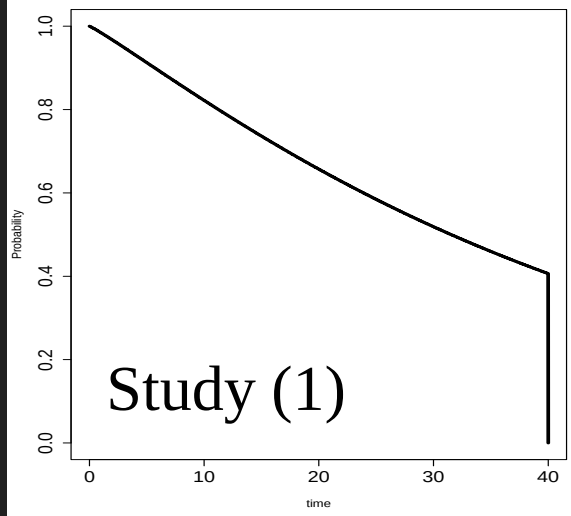
“TRUE” Survival functions

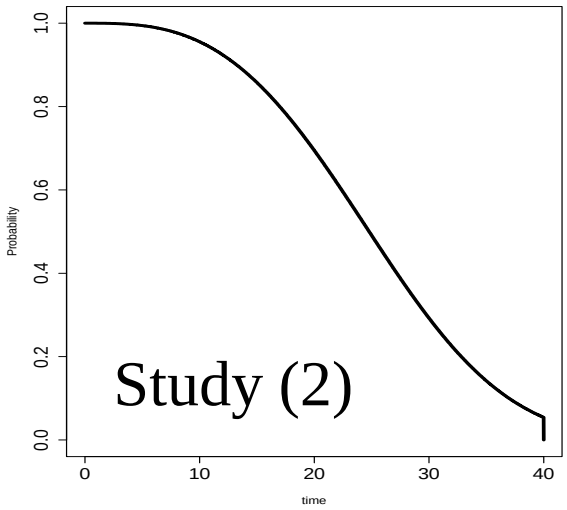


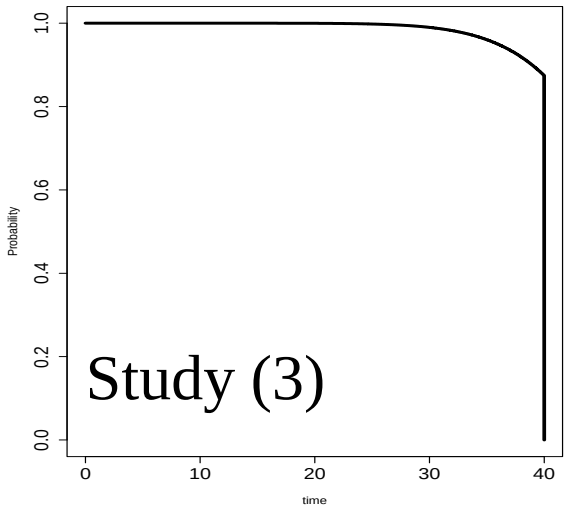
“TRUE” Hazard Ratio



Study-specific censoring distribution



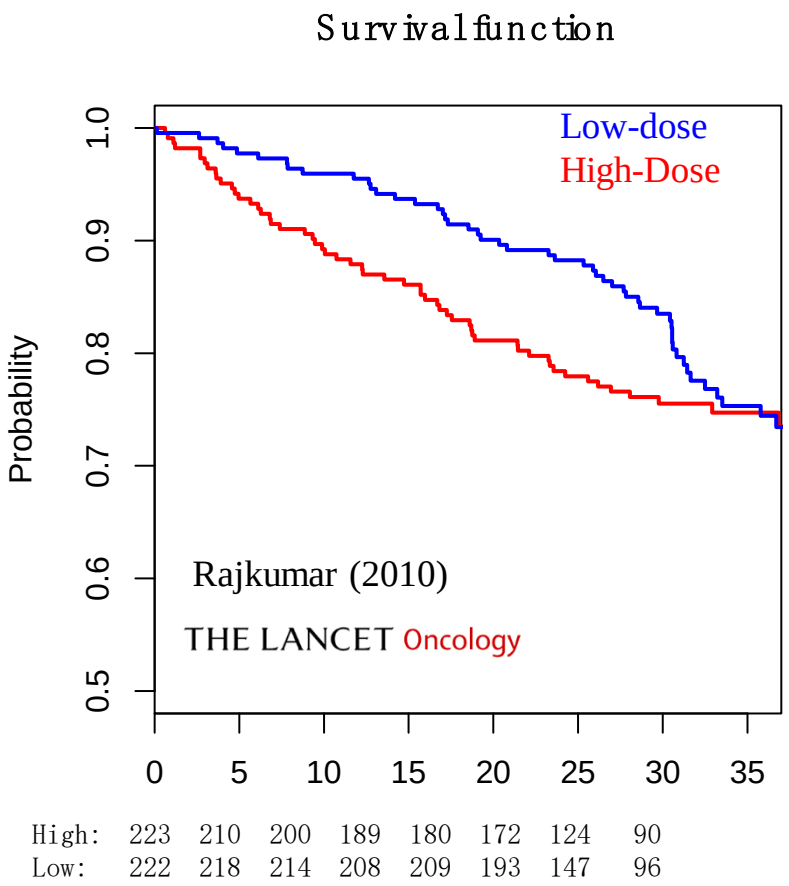


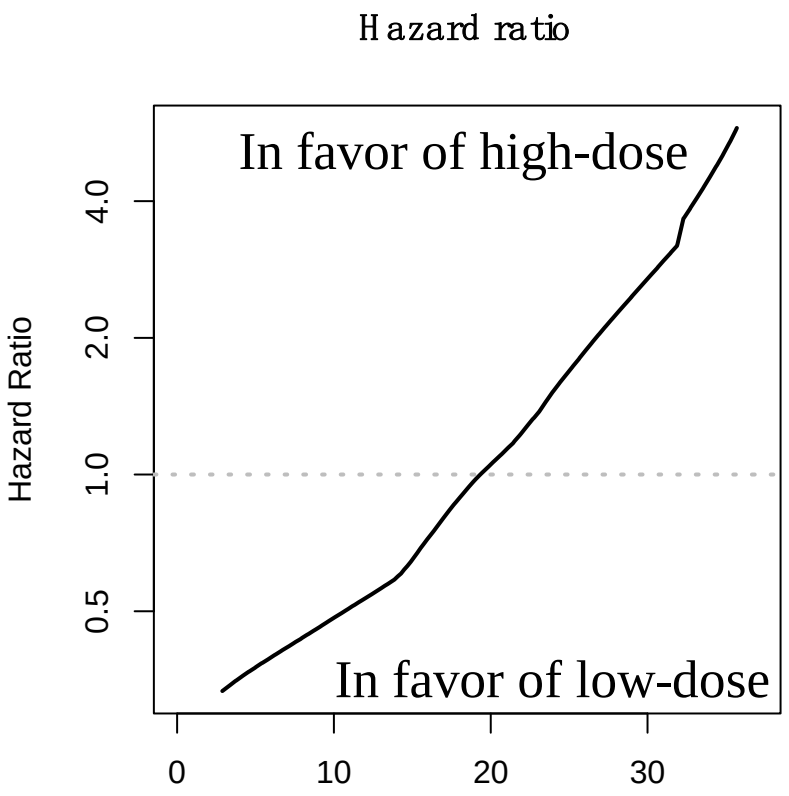


Example

- ECOG E4A03: A phase III randomized trial to compare low- and high-dose dexamethasone for newly diagnosed multiple myeloma
- N=445 (223 on high-dose, 222 on low-dose)
- One of the endpoints was overall survival

Example (ECOG E4A03)



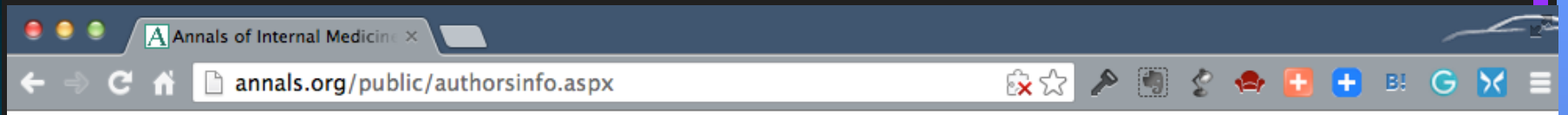


HR= 0.87 (0.95CI: 0.60 to 1.27), p=0.46

How do we interpret 0.87 ?

Ref. *Annals of Internal Medicine*, Guideline for Authors

http://annals.org/public/authorsinfo.aspx



Statistical Guidelines

Presentation	
Issue	Notes
Percentages	Report percentages to one decimal place (i.e., xx.x%) when sample size is >=200. To avoid the appearance of a level of precision that is not present with small samples, do not use decimal places (i.e., xx%, not xx.xx%) when sample size is < 200.
Standard	Use "mean (SD)" rather than "mean ± SD" notation. The ± symbol is ambiguous and can represent standard deviation or standard error.
Cox models	When reporting the findings from Cox proportional hazards models: ▪ Do not describe hazard ratios as relative risks. ▪ Do report how the assumption of proportional hazards was tested, and what the test showed.
P values	For *P* values between 0.001 and 0.20, please report the value to the nearest thousandth. For *P* values greater than 0.20, please report the value to the nearest hundredth. For *P* values less than 0.001, report as "*P*<0.001."
"Trend"	Use the word *trend* when describing a test for trend or dose-response. Avoid the term *trend* when referring to *P* values near but not below 0.05. In such instances, simply report a difference and the confidence interval of the difference (if appropriate) with or without the *P* value.
Statistical software	Specify in the statistical analysis section the statistical software—version, manufacturer, manufacturer's location, and the specific functions, procedures, or programs—used for analyses.
Cox models	When reporting the findings from Cox proportional hazards models: ▪ Do not describe hazard ratios as relative risks. ▪ Do report how the assumption of proportional hazards was tested, and what the test showed.

It seems checking the PH assumption is important …

- Check by your eye ball – (subjective…)
  - Log(-log(S(t))) vs. t
- Statistical tests
  - Include time-varying covariates in Cox's model
  - Goodness of fit tests
    - Schoenfeld residuals (Schoenfeld, 1982)
    - Weighted residuals (Grambsh & Therneau, 1994)
    - Cumulative residuals (Lin & Wei, 2002)

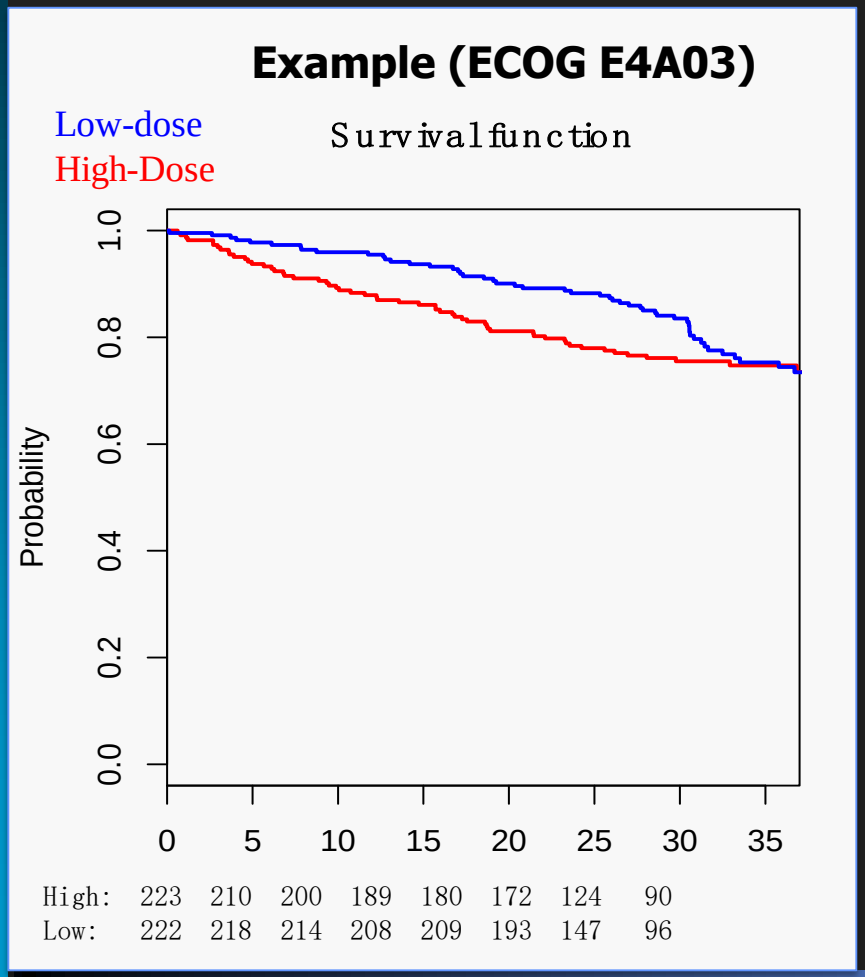
But, can we actually rule out non-PH cases by statistical tests?

Testing the PH assumption

- H_0: "PH is correct"
- H_1: "PH is NOT correct"

- Non-Significant test result does not necessarily means "H_0: PH is correct" is true
- Also, if sample size is huge, a significant p-value will be observed even with a tiny deviation from the PH

Need a single number to summarize the difference between two functions

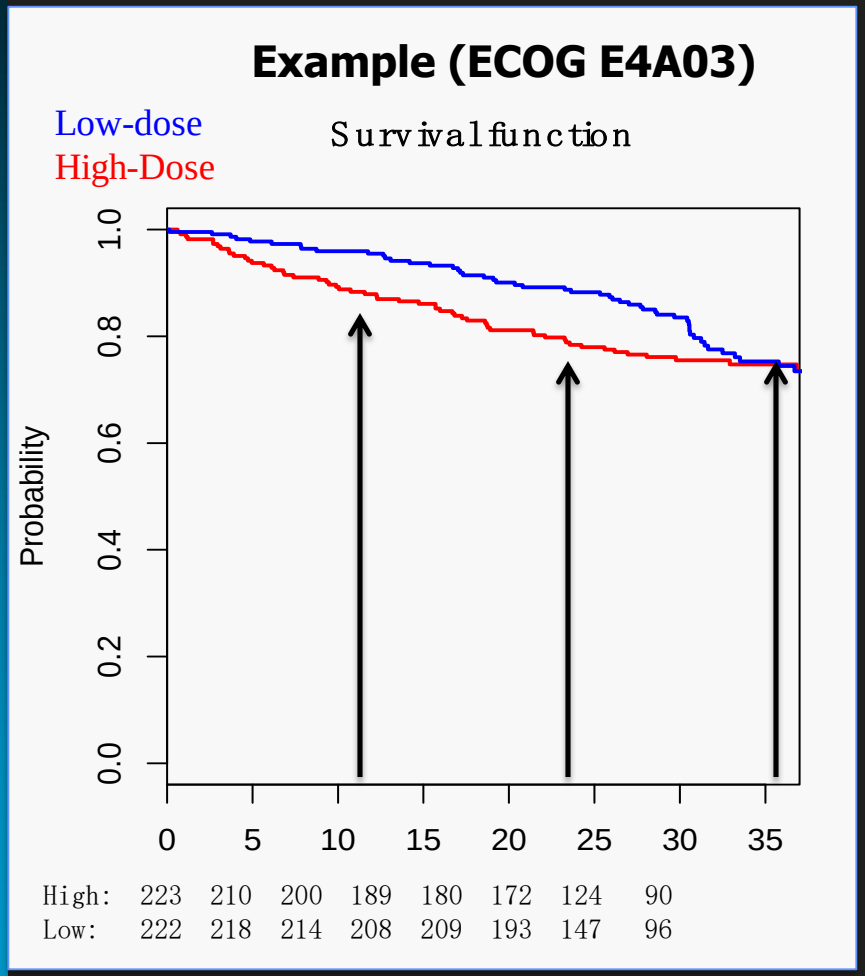


No perfect summary

Desirable ones:
... do not need a strong modeling assumption on the relationship between the two curves

→ model-free measures

(1) t-year survival probability

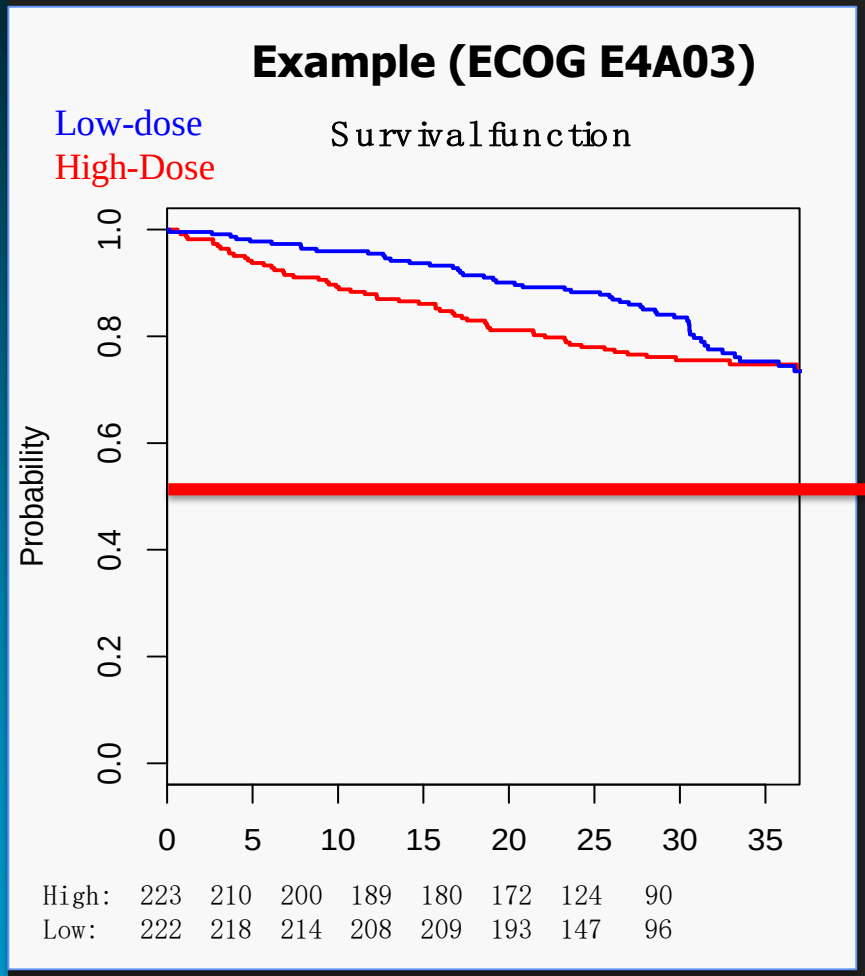


$$S_1(t) - S_0(t)$$

$$S_1(t)/S_0(t)$$

1-year?
2-year?
3-year?

(2) Median survival time



$$S_1^{-1}(0.5) - S_0^{-1}(0.5)$$

$$S_1^{-1}(0.5) \;/\; S_0^{-1}(0.5)$$

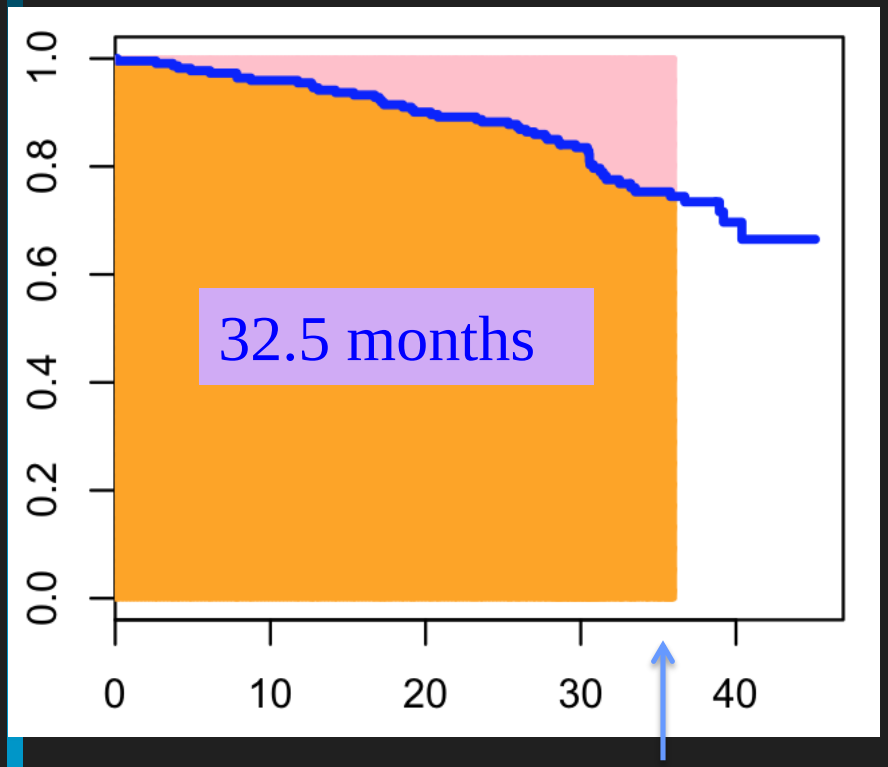
???

Sometimes this is inestimable…

(3) Restricted mean survival time (RMST)

$$\int_0^{\tau} S_1(u)du$$

Low-dose



Interpretation:

If you follow-up patients on low-dose for 36m, patients will survive 32.5 months *in average*

$(\tau-\text{year life expectancy})$

Note:

RMST is estimable even when median survival time is inestimable

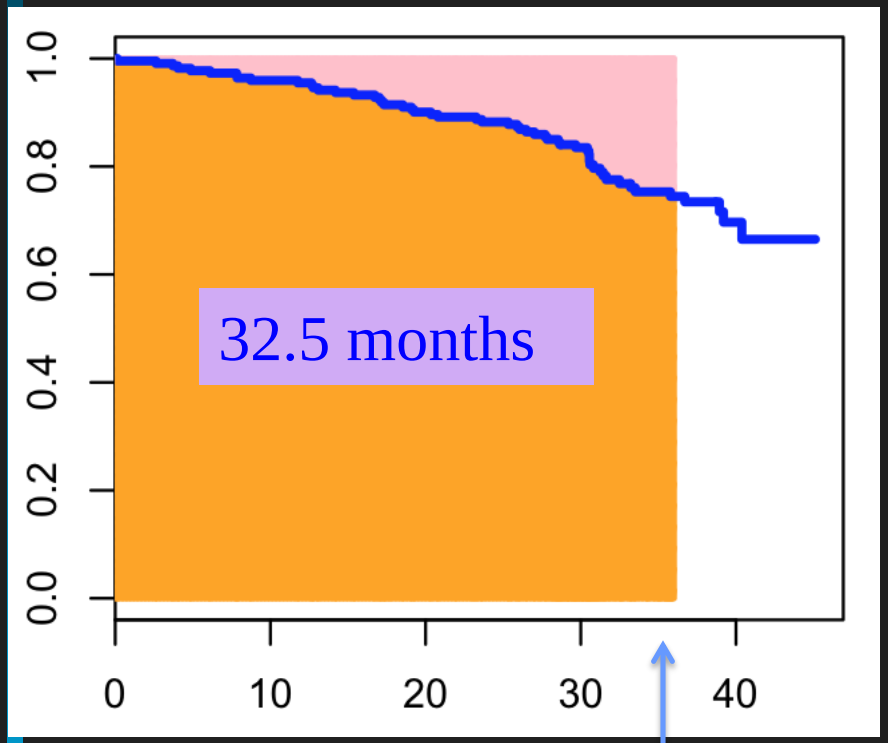
(3) Restricted mean survival time (RMST)

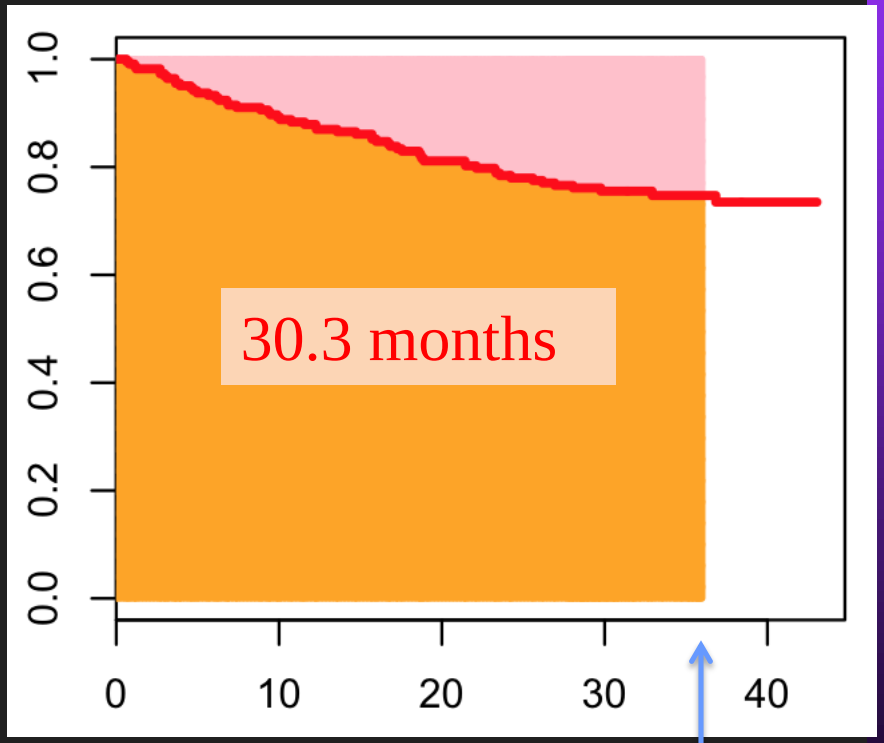
$$\int_0^{\tau} S_1(u)du$$

Low-dose

$$\int_0^{\tau} S_0(u)du$$

high-dose





$\tau = 36m$

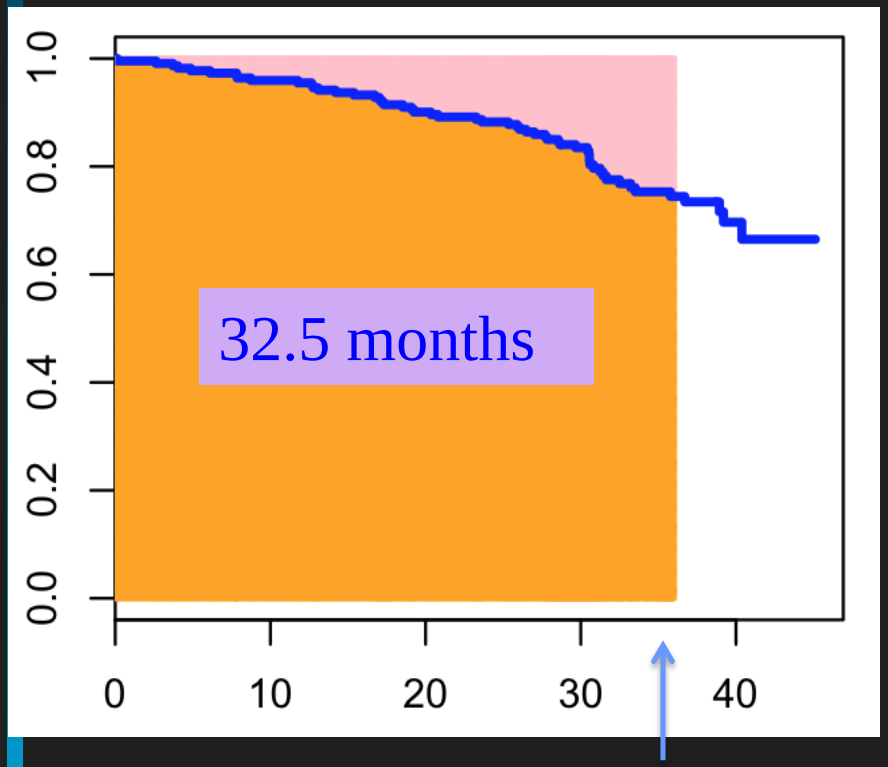
$\tau = 36m$

(3) Restricted mean survival time (RMST)

Difference in RMST: 2.2 months (0.95CI: 0.5 to 4.0, p<0.01)

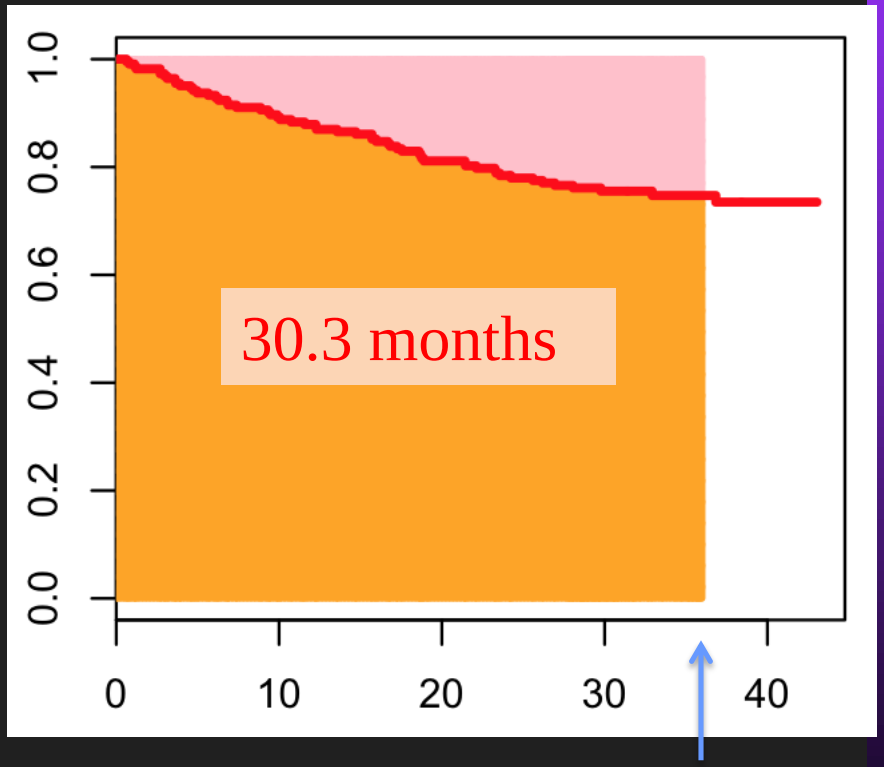
Recall: HR= 0.87 (0.95CI: 0.60 - 1.27), NonSignificant

Low-dose



$\tau = 36m$

high-dose

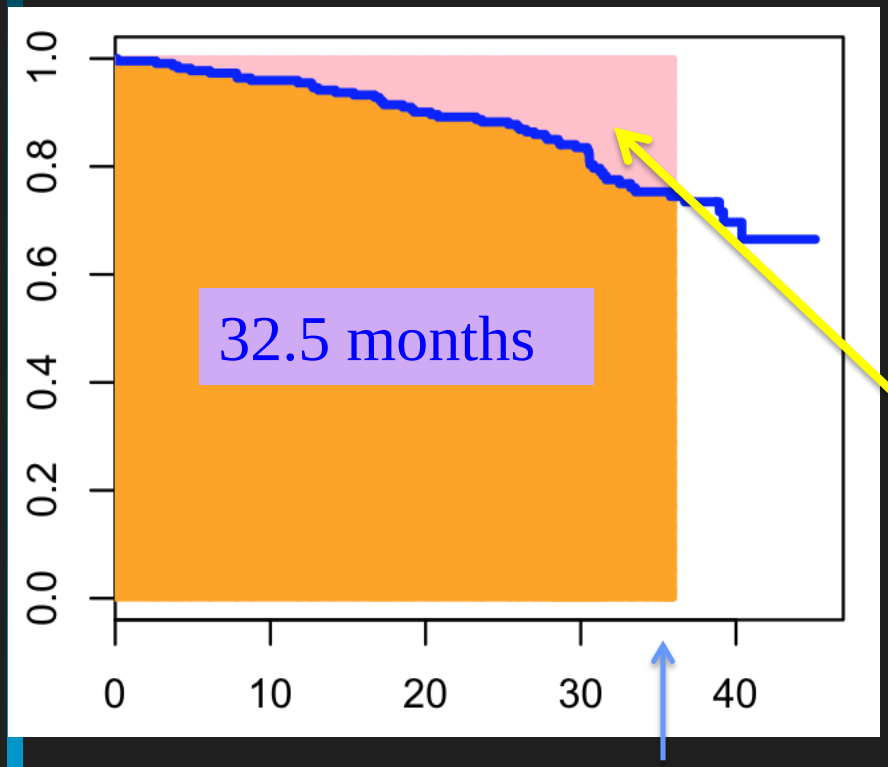


$\tau = 36m$

(4) Restricted mean time lost (RMTL)

$$\tau - \int_0^{\tau} S_1(u)du$$

Low-dose



The area **above** the survival curve

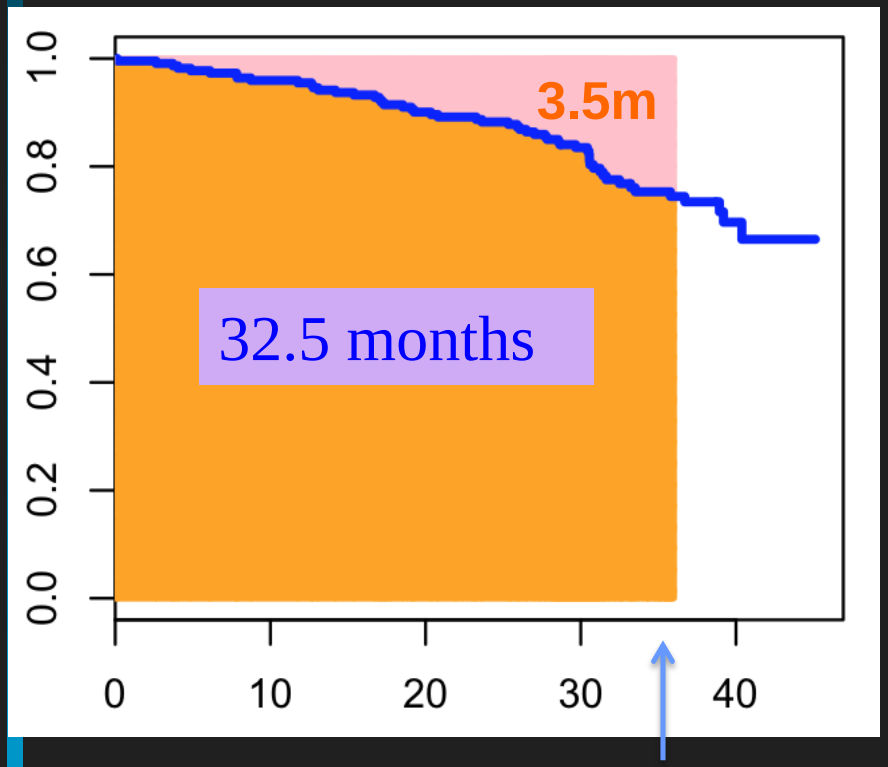
Interpretation

If you follow-up patients on low-dose for 36m, patients will lose time **3.5 months** *in average*

(4) Restricted mean time lost (RMTL)

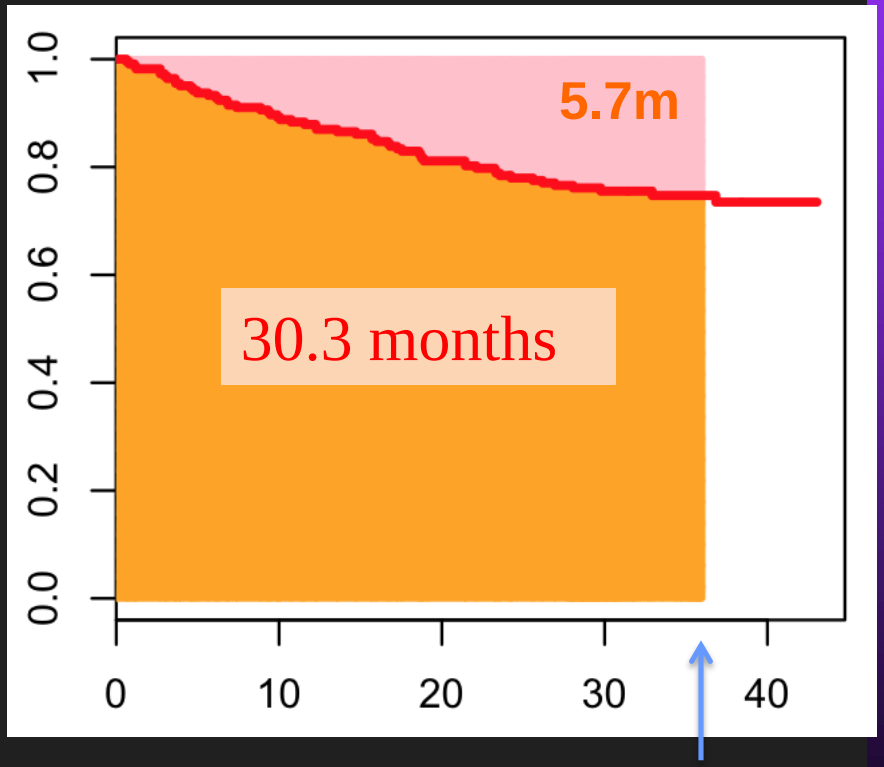
Ratio of RMTL: 3.5/5.7=0.61 (0.95CI: 0.42 to 0.90, p<0.01)

Low-dose



$\tau = 36m$

high-dose



$\tau = 36m$

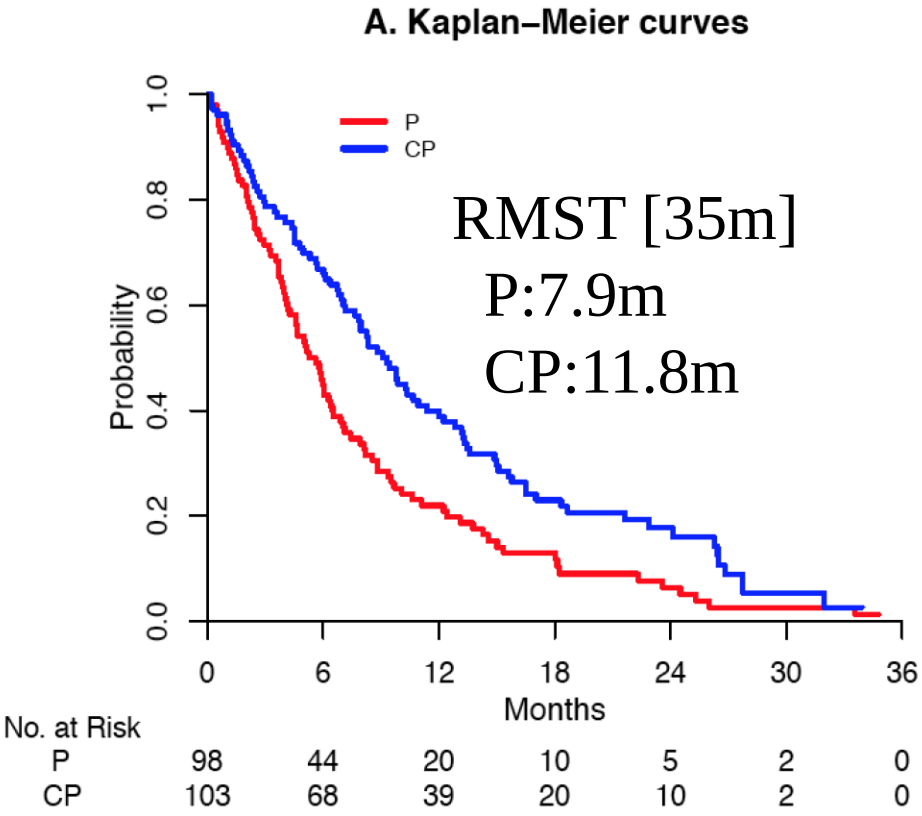
Lesson we learned from this example (a cross-hazards case)

- The PH assumption is clearly violated; the interpretation of HR is rather difficult
- Median survival time is a robust summary but was inestimable with this example
- RMST or RMTL would be a good alternative when summarizing the between-group difference
- HR-based tests did not have power (due to the cross-hazards), but a significant difference was seen with the RMST-based tests

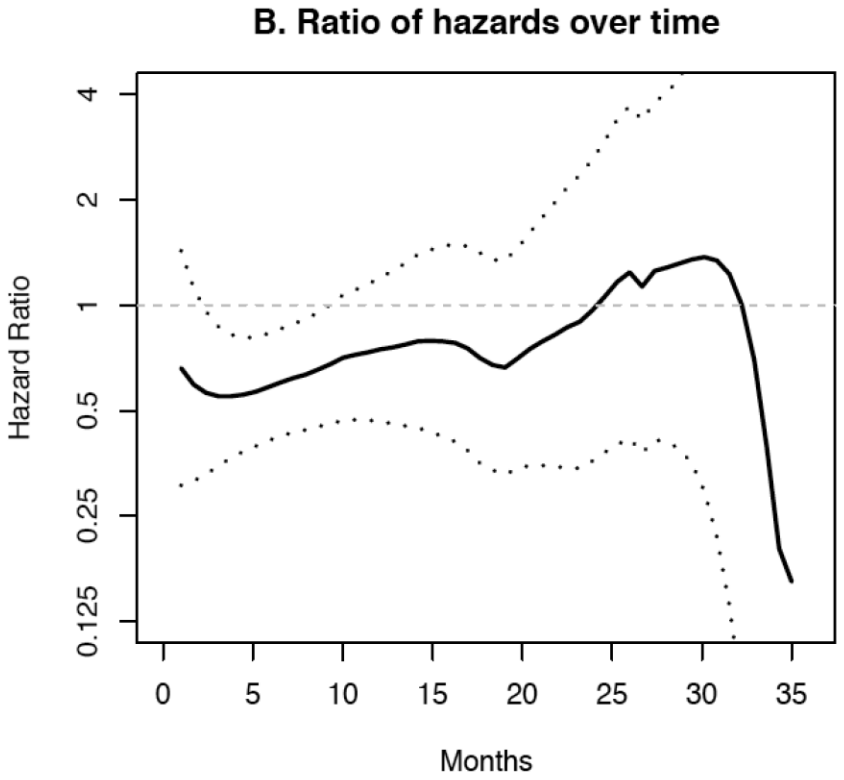
Other examples from cancer trials
Zukin et al. (2013, JCO)

- A randomized controlled trial to compare OS between single-agent pemetrexed (P) and the combination of carboplatin and pemetrexed (CP) in patients with advanced NSCLC with an ECOG PS of 2
- N=103 (CP arm) and N=98 (P arm)
- Median follow-up time is 27.5 months
- Median survival: CP (9.3m) and P (5.3m)

Zukin et al. (2013, JCO)
Advanced NSCLC, OS



RMST [35m]
P:7.9m
CP:11.8m



HR: 0.62 (0.46 to 0.83) **

RMST [35m] Difference:3.9 (1.5 to 6.3) **

RMST [35m] Ratio: 1.49 (1.17 to 1.91) **

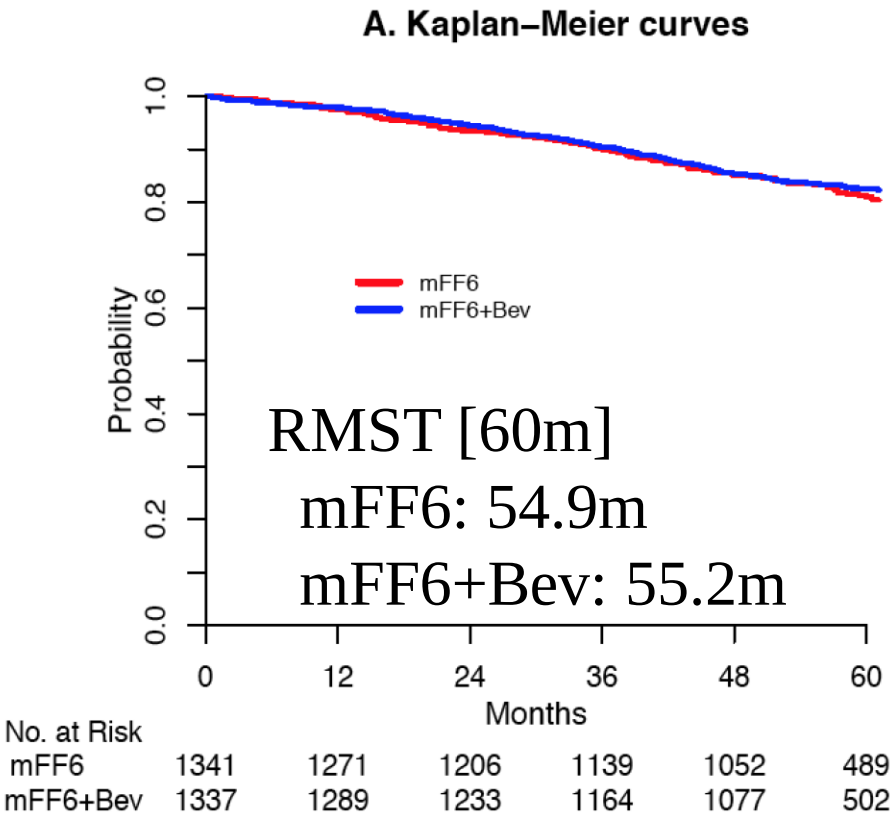
Lessons we learned from this example

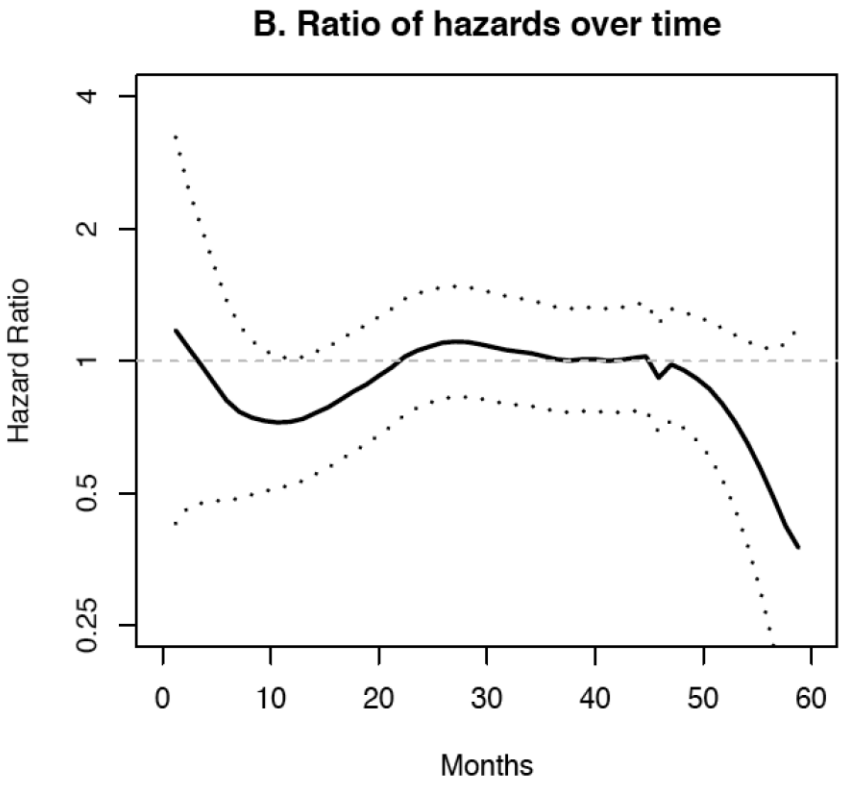
- RMST works for not only cross-hazard cases (“*qualitative interaction”)* but also non cross-hazard cases (“*quantitative interaction”)*
- The HR-based test also worked for this "*quantitative”* interaction example, but the resulting HR estimate would be still hard to interpret

Allegra et al. (2013, JCO)

- A randomized trial to assess efficacy and safety of the combination of bevacizumab and modified FOLFOX6 (mFF6), as the adjuvant therapy, in patients with stage II/III colon cancer.
- N=1341 (mFF6 arm) and N=1337 (bev+mFF6 arm)
- Median follow-up time is 4.9 years
- Median survival: Not reached

Allegra et al. (2013, JCO)
Stage II/III colon cancer, OS





HR: 0.95 (0.79 to 1.13)

RMST [60m] Difference: 0.3 (-0.7 to 1.3) months

RMST [60m] Ratio: 1.00 (0.99 to 1.02)

Lessons we learned from this example

- The confidence interval for the hazard ratio is rather wide (0.79 to 1.13) due to relatively small number of events. The evidence from this analysis is inconclusive due to the lack of information

- The CI for difference (or ratio) of RMSTs are tight around 0 (or 1). We can claim that (at least for the RMST) these two groups are not *clinically* significantly different (<2% difference over 60 months)

Summary

- Need to be cautious when using model-based metrics, such as HR
  - When model assumptions do not hold, the interpretation is rather difficult
- Model-free and clinically interpretable metrics such as RMST would be a better choice, unless we are confident that the model assumption is correct

Implementations in R

survRM2 package

- survRM2_1.0.tar.gz (Source)
- survRM2_1.0.tgz (Mac binary)
- survRM2_1.0.zip (Windows binary)
- survRM2_vignette (Package Vignette)

What the package can do:

- Calculating RMST and CI for each group
- Two-sample comparison based on the RMST (difference in RMST, ratio of RMST, and ratio of RMTL)
- ANCOVA type covariate adjustment for these three between-group contrast measures

Sample data used in the package manual/vignette (PBC data, Fleming & Harrington 1991)

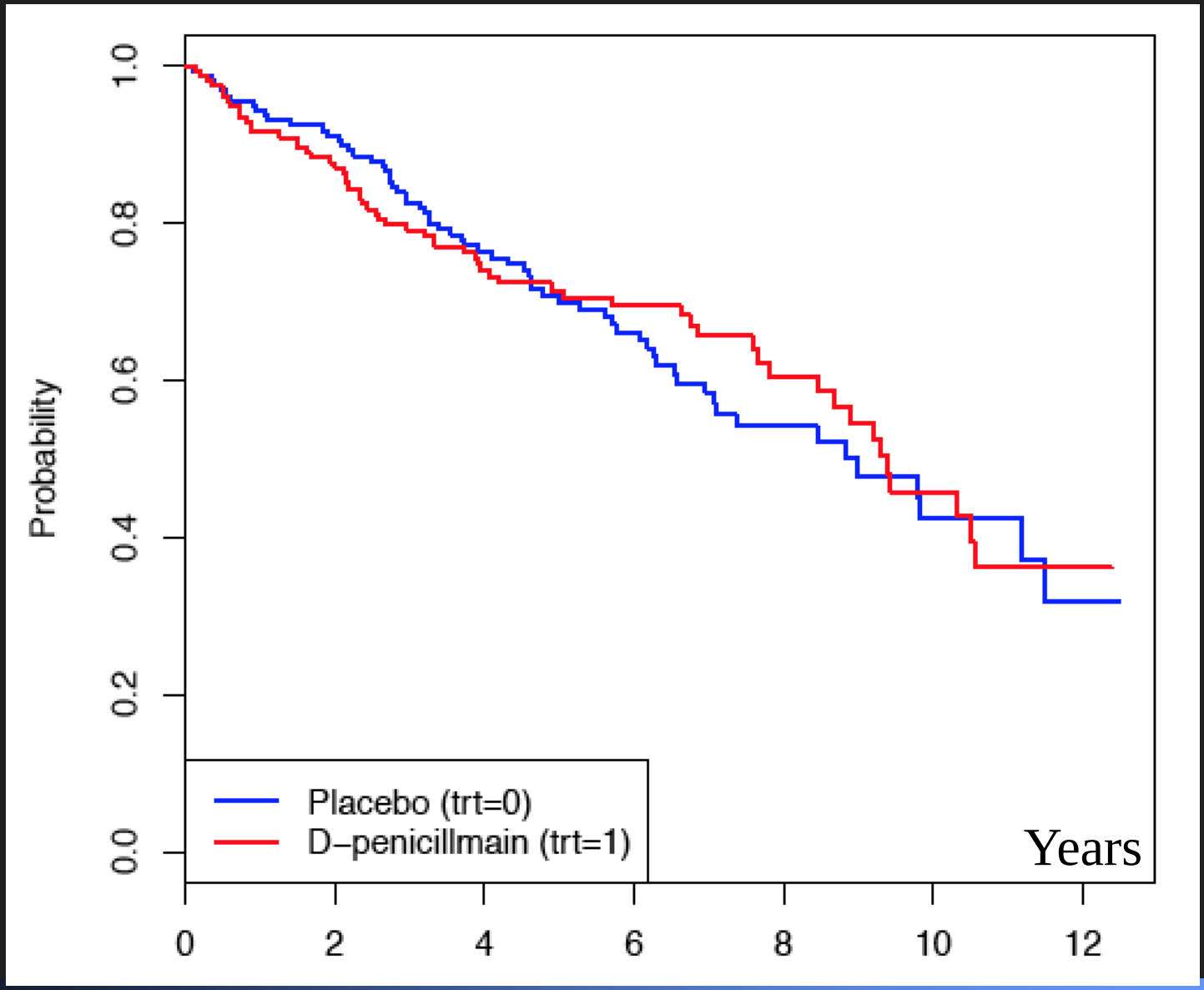
```
> library(survRM2)
> D=rmst2.sample.data()

> nrow(D)
[1] 312

> head(D)
```

	time	status	arm	age	edema	bili	albumin	protime
1	1.095140	1	0	58.76523	1.0	14.5	2.60	12.2
2	12.320329	0	0	56.44627	0.0	1.1	4.14	10.6
3	2.770705	1	0	70.07255	0.5	1.4	3.48	12.0
4	5.270363	1	0	54.74059	0.5	1.8	2.54	10.3
5	4.117728	0	1	38.10541	0.0	3.4	3.53	10.9
6	6.852841	1	1	66.25873	0.0	0.8	3.98	11.0

KM plots


Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
12
Years
Placebo (trt=0)
D−penicillmain (trt=1)

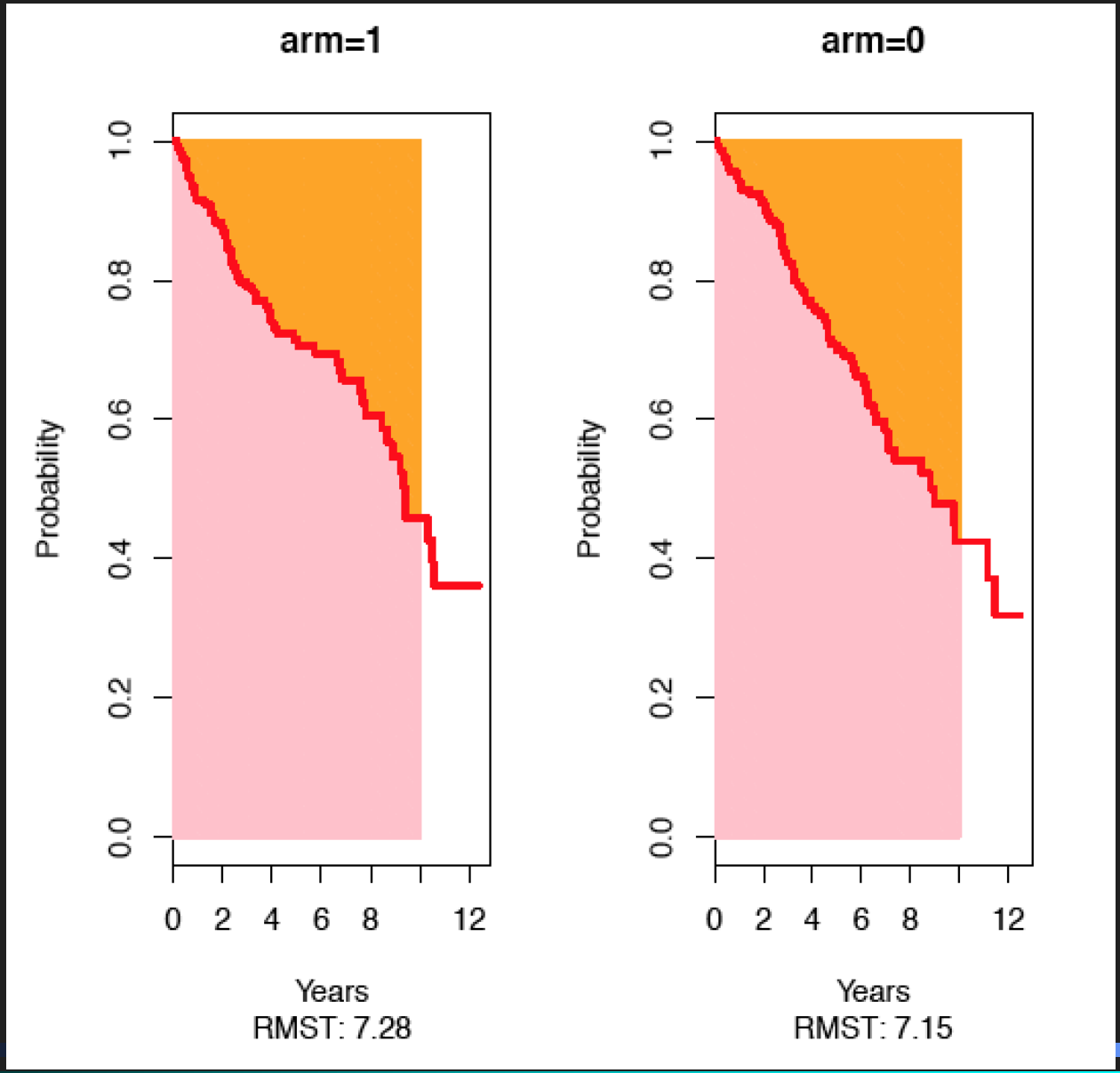
```
> foo=rmst2(time, status, arm, tau=10)
> print(foo)
The truncation time: tau = 10 was specified.

Restricted Mean Survival Time (RMST) by arm
               Est.      se   lower .95   upper .95
RMST (arm=1) 7.283   0.295       6.704       7.863
RMST (arm=0) 7.146   0.283       6.592       7.701

Restricted Mean Time Lost (RMTL) by arm
               Est.      se   lower .95   upper .95
RMLT (arm=1) 2.717   0.295       2.137       3.296
RMTL (arm=0) 2.854   0.283       2.299       3.408

Between-group contrast
                      Est. lower .95 upper .95       p
RMST (arm=1)-(arm=0) 0.137    -0.665     0.939   0.738
RMST (arm=1)/(arm=0) 1.019     0.912     1.139   0.738
RMTL (arm=1)/(arm=0) 0.952     0.714     1.270   0.738
```

```
> plot(foo, xlab="Years", ylab="Probability"))
```



About the truncation time “tau”

Default value:

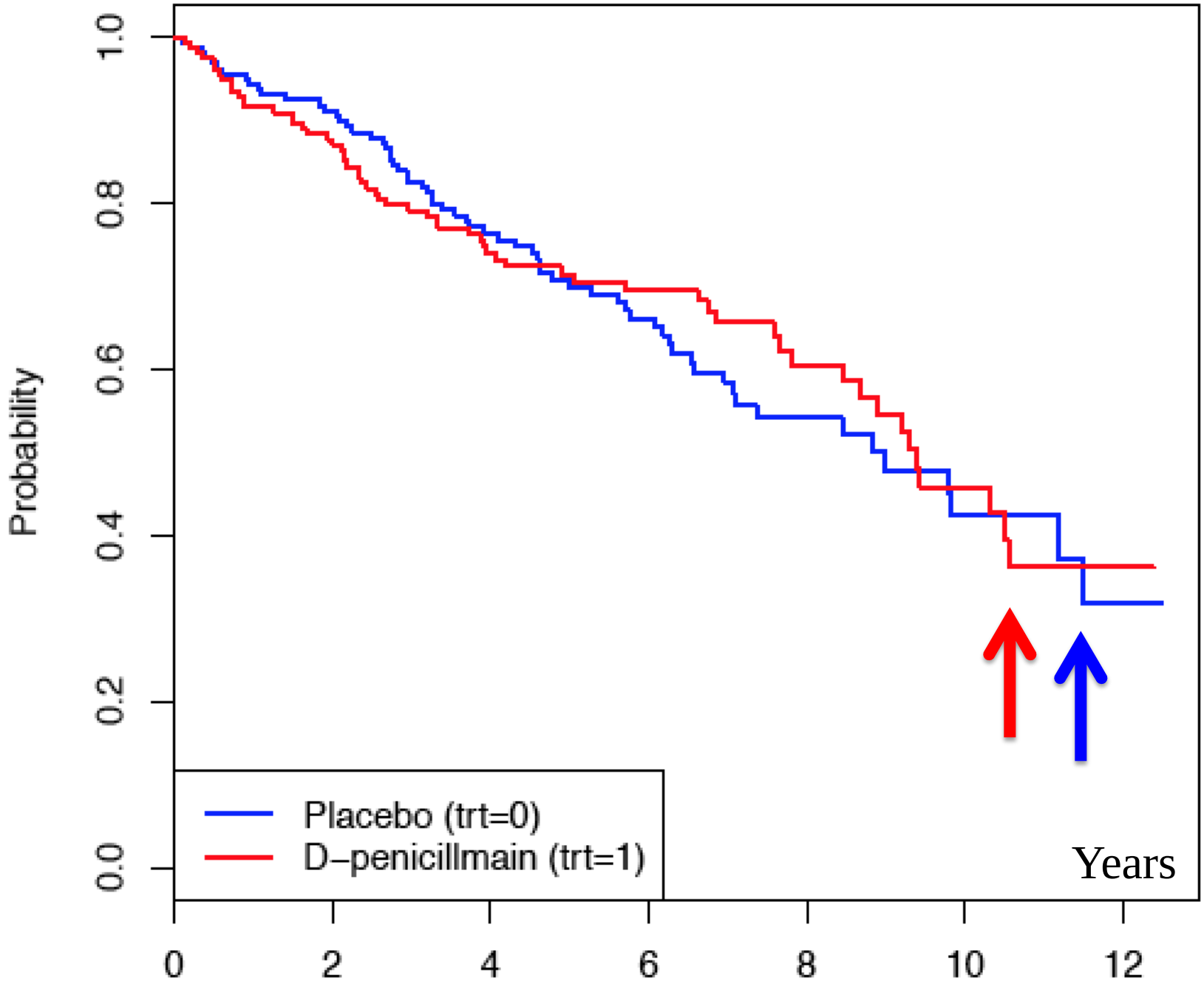
When tau is NULL or not specified in rmst2()

```
> rmst2(time, status, arm, tau=NULL)
```

or

```
> rmst2(time, status, arm)
```

The default tau is the minimum of the largest observed event time in each of the two groups


Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
12
Years
Placebo (trt=0)
D−penicillmain (trt=1)

About the truncation time "tau"

In RCT, tau should be pre-specified in the protocol

Note that *tau* needs to be smaller than or equal to the default *tau*, because the KM cannot well estimate the survival function beyond the last observed event time

In this example, the default tau was 10.5 years

```
> rmst2(time, status, arm, tau=12)

Error in rmst2(time, status, arm, tau = 12) :
The truncation time, tau, needs to be shorter than or equal to
the minimum of the largest observed event time on each of the
two groups:  10.549
```

Adjusted analysis

- The package also implements ANCOVA-type adjusted analyses proposed by Tian et al. (2014, Biostatistics)

```
> head(x)
       age   bili albumin
1 58.76523 14.5     2.60
2 56.44627  1.1     4.14
3 70.07255  1.4     3.48
4 54.74059  1.8     2.54
5 38.10541  3.4     3.53
6 66.25873  0.8     3.98

> rmst2(time, status, arm, tau=10, covariates=x)
```

```
> rmst2(time, status, arm, tau=10, covariates=x)

The truncation time: tau = 10  was specified.

Summary of between-group contrast (adjusted for the covariates)
                       Est. lower .95 upper .95     p
RMST (arm=1)-(arm=0) 0.210    -0.463     0.883 0.540
RMST (arm=1)/(arm=0) 1.033     0.937     1.140 0.514
RMTL (arm=1)/(arm=0) 0.966     0.752     1.240 0.786


Model summary (difference of RMST)
             coef se(coef)      z     p lower .95 upper .95
intercept  2.533    2.090  1.212 0.226    -1.563     6.629
arm        0.210    0.343  0.613 0.540    -0.463     0.883
age       -0.069    0.018 -3.900 0.000    -0.103    -0.034
bili      -0.325    0.039 -8.386 0.000    -0.401    -0.249
albumin    2.550    0.472  5.401 0.000     1.624     3.475
```

Model summary (ratio of RMST)

	coef	se(coef)	z	p	exp(coef)	lower .95	upper .95
intercept	1.336	0.348	3.834	0.000	3.803	1.921	7.529
arm	0.033	0.050	0.652	0.514	1.033	0.937	1.140
age	-0.009	0.003	-3.410	0.001	0.991	0.985	0.996
bili	-0.087	0.013	-6.523	0.000	0.917	0.893	0.941
albumin	0.360	0.080	4.491	0.000	1.434	1.225	1.678

Model summary (ratio of time-lost)

	coef	se(coef)	z	p	exp(coef)	lower .95	upper .95
intercept	2.027	0.689	2.943	0.003	7.591	1.968	29.274
arm	-0.035	0.127	-0.272	0.786	0.966	0.752	1.240
age	0.025	0.007	3.810	0.000	1.026	1.012	1.039
bili	0.063	0.008	8.334	0.000	1.065	1.049	1.080
albumin	-0.750	0.149	-5.033	0.000	0.472	0.353	0.633

nature aging

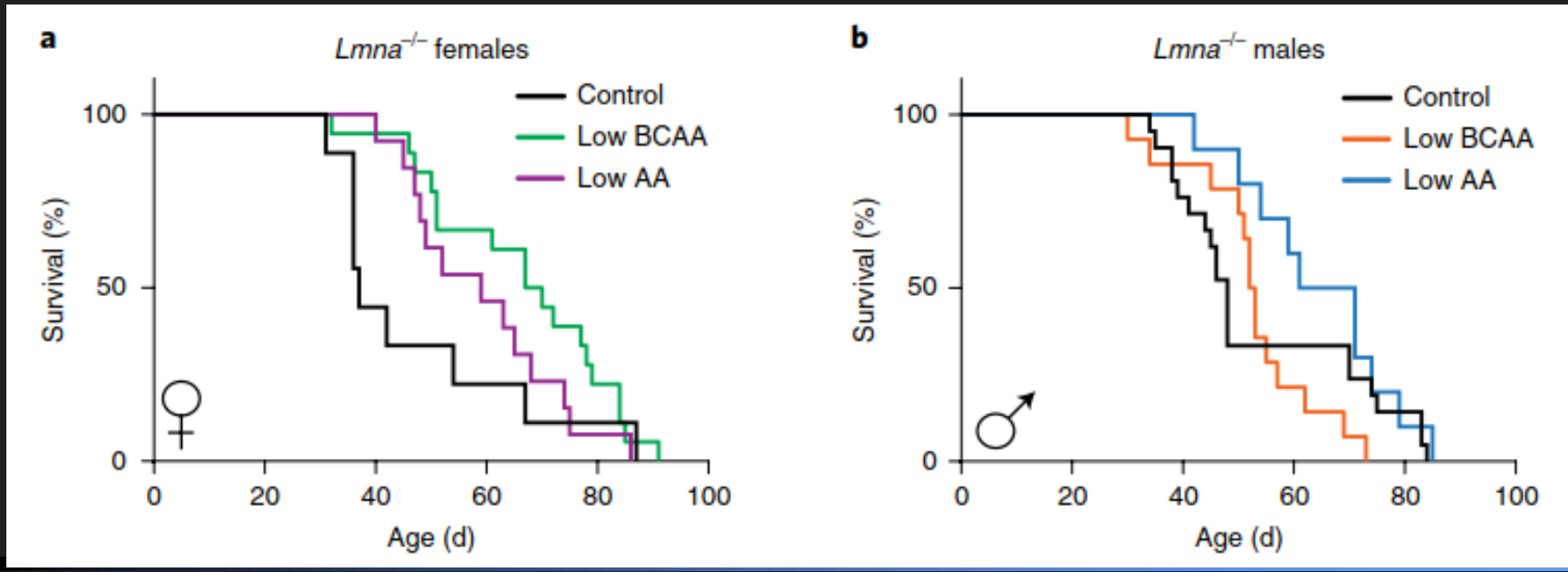
ARTICLES

https://doi.org/10.1038/s43587-020-00006-2

Check for updates

Lifelong restriction of dietary branched-chain amino acids has sex-specific benefits for frailty and life span in mice

Nicole E. Richardson[1,2,3], Elizabeth N. Konon[1,2], Haley S. Schuster[1,2], Alexis T. Mitchell[1,2],



Conclusion from the paper

- The results demonstrate that restricting dietary branched-chain amino acids (BCAAs) can increase health span and longevity in mice and suggest that reducing dietary BCAAs may hold potential as a translatable intervention to promote healthy aging.
- They claimed that low BCAA-fed $Lmna^{-/-}$ female mice lived longer than the control group (P=0.0794, log-rank test; P=0.0098, Wilcoxon test), and low AA-fed $Lmna^{-/-}$ female mice also lived longer than those with the control diet (P=0.366, log-rank test; P=0.0355, Wilcoxon test).

Weighted Log-rank Test

- The Wilcoxon test is more sensitive to the early differences of survival curves by assigning more weights to early follow-ups (the log-rank test uses an equal weight of 1), and thus the convergence pattern towards day 85 is downplayed.

RMST

Table·1·The·restricted·mean·survival·time·(RMST)†·of·the·low·branched-chain·amino·acids·(BCAA)·and·low·amino·acids·(AA)·and·control·groups·based·on·the·original·data*·for·lamin·A/C-deficient·(*Lmna*$^{-/-}$)·female·mice.·¶

Group¤	RMST¤	95%·CI¤	RMST·Difference·vs.·Control¤	P-value¤
Low·BCAA¤	66.2¤	[58.7,·73.7]¤	18.9·[5.1,·32.6]¤	0.007¤
Low·AA¤	59.3¤	[52.1,·66.6]¤	12.0·[-1.6,·25.6]¤	0.084¤
Control¤	47.3¤	[35.8,·58.8]¤	\¤	\¤

*Data·were·obtained·from·the·Supplementary·Table·2·in·the·original·paper[1].·¶

†The·restricted·mean·survival·time·(RMST)·was·estimated·by·calculating·the·area·under·the·survival·curve·using·the·"survRM2"·package·in·R·software,·version·3.6.0·(R·Project·for·Statistical·Computing).¶

Conclusions

- Hazard ratio estimate and log-rank test are routinely used for designing, monitoring and analyzing clinical studies with time to event data, but it is not always the best approach
- When there is no strong empirical or biological evidence that the PH assumption is valid, a model-free summary measure for the between-group difference is preferred
- We should not tend to stay in our comfort zone. We should be brave enough to employ alternative methods in practice, when they are more appropriate