Biostatistics

Mithat Gonen

Brenden Bready

New Material from Last Week

- We discussed “sequential” or “group sequential” experiments
  - You do one or more interim analyses to see if you can stop the experiment early
  - If “significant” stop and reject the null, if not continue
- Example: Your total sample size is 10 per group, you have two groups and you do one interim analysis when you reach 5 per group
  - If you use $p<0.05$ both at interim analysis and at the end your Type I error probability with be 0.075.

How to do it Right?

- There are many (inifinetely many) pairs of p-values thresholds (one for interim one for final) that will make it work
  - If you use $p<0.03$ both at the interim analysis and at the final analysis, then your Type I error probability is 0.05.
  - If you use $p<0.009$ at the interim analysis and $p<0.0467$ at the final analysis, then your Type I error probability is 0.05.

Another Example

- Optimal Cutoff
- A continuous predictor, either a continuous, binary or survival outcome
- Goal is to find the optimal cutoff to dichotomize the predictor and then a generate a p-value for the resulting two sample comparison
- If you use $p<0.05$ your Type I error is probability 0.40 !!!
- You need a tailored method for this (called maximal chi-square method)

Analysis of Binary Outcomes

- We have two groups and a binary outcome
- Ex: Treatment and response
- N_1 patients in treatment A and N_2 in treatment B
- R_1 responders (of N_1) in A and R_2 (of N_2) in B
- Response rates of R_1/N_1 and R_2/N_2

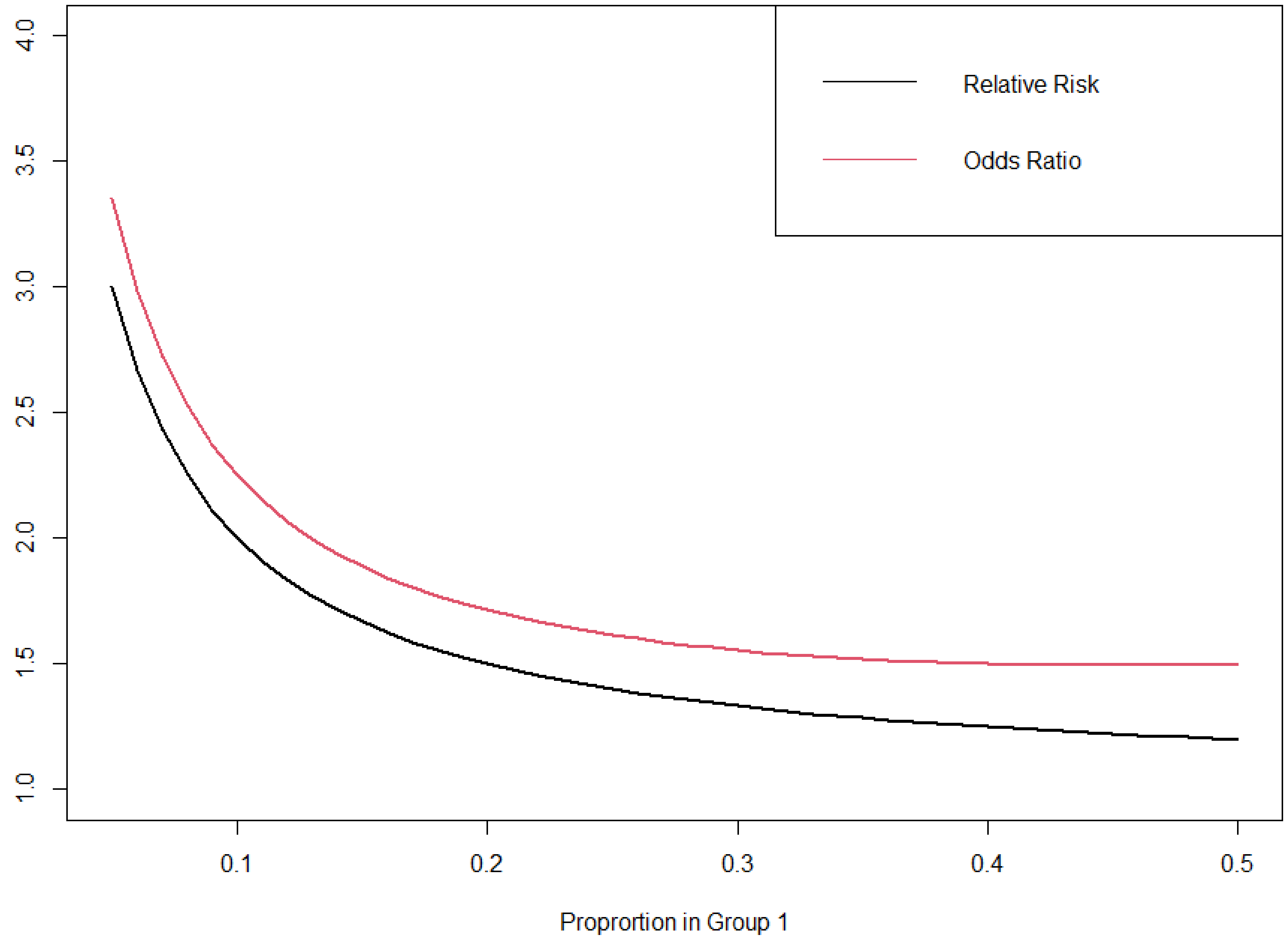
2x2 Table

	Response	No Response	TOTAL
Treatment A	R_1	$N_1 - R_1$	N_1
Treatment B	R_2	$N_2 - R_2$	N_2
TOTAL	$R_1 + R_2$	$N_1 - R_1 + N_2 - R_2$	$N_1 + N_2$

- Think of a 2x2 table anytime you have a binary outcome and 2 groups
- Think of a KxP table any time you have an outcome with K categories and P groups
- All the information in the data we need is here
- Very useful and flexible display

Difference, Relative Risk, Odds Ratio

- Proportions: R1/N1 = p1 and R2/N2 = p2
- Three metrics for summarizing the difference
  - Difference: p1 – p2
  - Relative Risk: p1/p2
  - Odds Ratio: [p1/(1-p1)] / [p2/(1-p2)]
- Important note: Odds Ratio is often interpreted as a relative risk (e.g., 2-fold increase in risk) but it is not



Relative Risk
Odds Ratio
1.0
1.5
2.0
2.5
3.0
3.5
4.0
0.1
0.2
0.3
0.4
0.5
Proprortion in Group 1

Hypothesis Testing in a 2x2 Table

	Response	No Response	TOTAL
Treatment A	R_1	$N_1 - R_1$	N_1
Treatment B	R_2	$N_2 - R_2$	N_2
TOTAL	$R_1 + R_2$	$N_1 - R_1 + N_2 - R_2$	$N_1 + N_2$

- Null hypotheses: Response rates are the same
- Alternative: They are not the same
  - Is this one-sided or two-sided?
- How to test this null?
  - We need a test statistic

Chi-Square Test

- For historical reasons, the derivation of this test was done differently and involved the chi-square distribution instead of the normal
- This idea more easily generalizes to bigger tables
- Chi-square test is always two-sided (memorize this) and does not lend itself to a confidence interval easily
- But used all the time in practice, partly because it applies to difference, relative risk and odds ratio equally.

Fisher's exact test

	Response	No Response	TOTAL
Treatment A	R_1	$N_1 - R_1$	N_1
Treatment B	R_2	$N_2 - R_2$	N_2
TOTAL	$R_1 + R_2$	$N_1 - R_1 + N_2 - R_2$	$N_1 + N_2$

- Under the null hypothesis of equal response rates and N_1 patients in treatment A and N_2 in B, how many do you expect to response in Treatment A?
- Involves something called a hypergeometric distribution
- Use it if available, definitely use it in small tables (even if one cell is small; eye of the beholder but <5 is commonly cited)

Summary of the 2x2 Table

- Three metrics: Difference, Relative Risk, Odds Ratio
- Two test statistics
  - Chi-Square test
  - Fisher's exact test

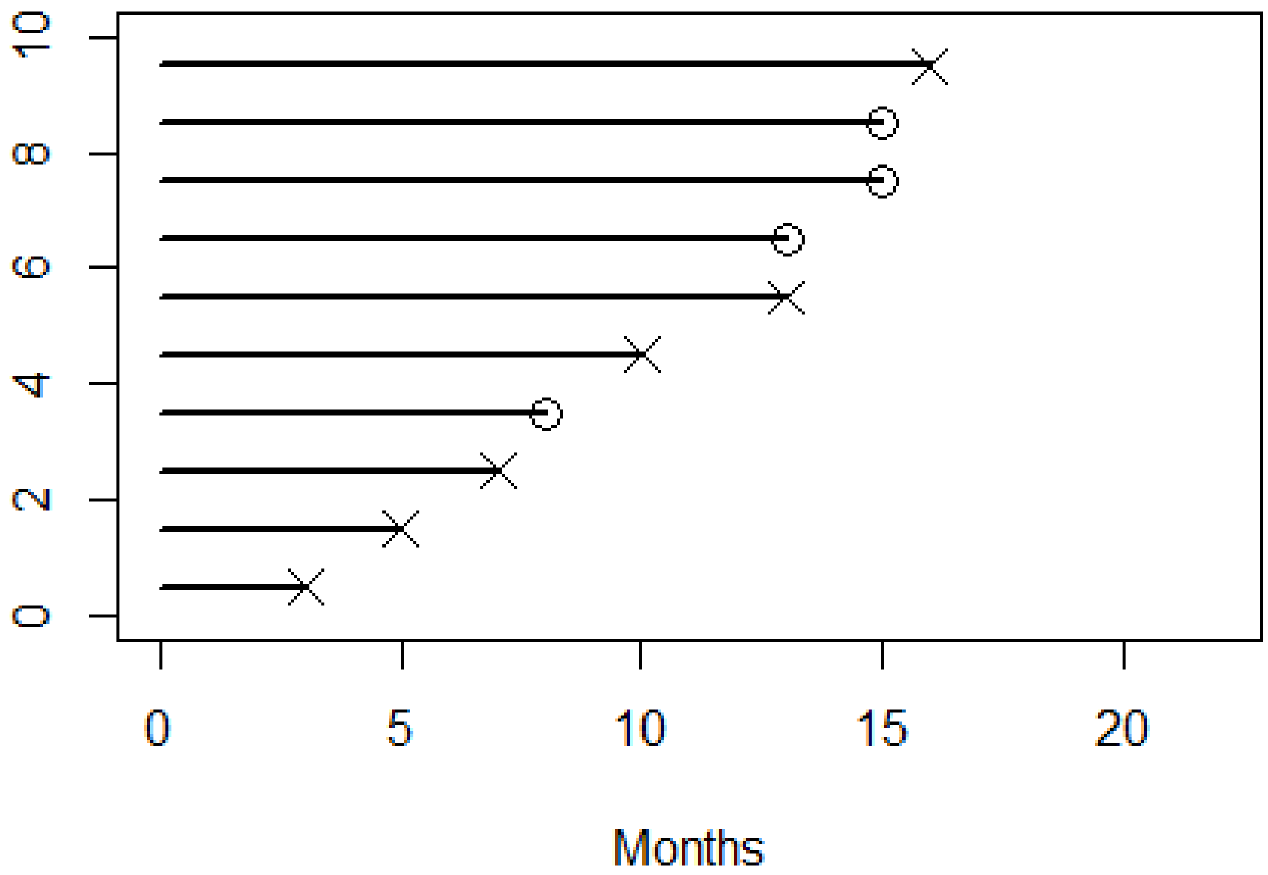
Survival Analysis

- Most important cancer outcomes are time-to-event
  - Time to death (overall survival)
  - Time to cancer death (disease specific survival)
  - Time to progression (progression free survival)
- Time interval between a “baseline” and an event
- Continuous variable?
- What is the big deal?

Censoring

- You are VERY VERY unlikely to observe all patients have the event of interest
- Some people will be alive or free of progression when the data analysis time comes
- This is called censoring, a data point that is not (yet) fully observed
  - Unfortunate terminology
- Primary reason why survival analysis requires special treatment

Data:
3, 5, 7, 8+, 10,
13, 13+, 15+,
15+, 16



What Can We Do?

- Ignore censoring, analyze as continuous
  - Treats people under follow-up as if they died
  - Underestimates survival probabilities
- Ignore time, analyze binary
  - Someone who died at 10 years vs someone who is alive and under follow-up at 6 months
- Both wrong, should NEVER EVER be used

Toy Example

- Data: 3, 5, 7, 8+, 10, 13, 13+, 15+, 15+, 16
- Do you remember the cumulative distribution function (CDF)?
- Plot P(survival time <= t) against t
- We want the survival function: P(survival time > t) = 1 - CDF
- Let’s ignore censoring for a moment and act if everyone is dead
- P(survival time > t) = 1 until t = 3
- P(survival time > t) = 9/10 = 0.1 for t between 3 and 5
- P(survival time > t) = 8/10 = 0.2 for t between 5 and 7
- (# alive before or on t)/n

Another way to calculate the survival function

- $S(t) = 1$, until $t = 3$
- $S(t) = 9/10 = 0.9$, between $t=3$ and $t=5$
- $S(t) = (9/10)*(8/9) = 8/10 = 0.8$

Probability of making it to 3 months

Probability of making it to 5 months, given you made it to 3 months

9 people made it to 3 months, 8 of these 9 made it 5 months

New denominator: 9 people at risk of death at 5 months

9 People Are "AT RISK" at 3 months

Next time point t = 7

- Prob of making it to 5 months = ?

3, 5, 7, 8+, 10, 13, 13+, 15+, 15+, 16

Next time point t = 7

- Prob of making it to 5 months = (8/10)
- Prob of making to 7 months given you have made it to 5 months = ?

3, 5, 7, 8+, 10, 13, 13+, 15+, 15+, 16

Next time point t = 7

- Prob of making it to 5 months = (8/10)
- Prob of making to 7 months given you have made it to 5 months = (7/8)
- Prob of making to 7 months = (8/10)*(7/8) = (7/10) = 0.7

3, 5, 7, 8+, 10, 13, 13+, 15+, 15+, 16

Next time point t = 8+

- Prob of making it to 7 months = ?

3, 5, 7, 8+, 10, 13, 13+, 15+, 15+, 16

Next time point t = 8+

- Prob of making it to 7 months = (7/10)
- Prob of making to 8 months given you have made it to 7 months = ?

3, 5, 7, 8+, 10, 13, 13+, 15+, 15+, 16

Next time point t = 8+

- Prob of making it to 7 months = (7/10)
- Prob of making to 8 months given you have made it to 7 months = (7/7)
- Prob of making to 7 months = (7/10)*(7/7) = (7/10) = 0.7

3, 5, 7, 8+, 10, 13, 13+, 15+, 15+, 16

Next time point t = 10

- Prob of making it to 8 months = (7/10)
- Prob of making to 10 months given you have made it to 8 months = ?

3, 5, 7, 8+, 10, 13, 13+, 15+, 15+, 16

Next time point t = 10

- Prob of making it to 8 months = (7/10)
- Prob of making to 10 months given you have made it to 8 months = ?
- How many people are at risk past 8 months?

3, 5, 7, 8+, 10, 13, 13+, 15+, 15+, 16

Next time point t = 10

- Prob of making it to 8 months = (7/10)
- Prob of making to 10 months given you have made it to 8 months = (5/6)
- Prob of making to 7 months = (7/10)*(5/6) = 0.583

3, 5, 7, 8+, 10, 13, 13+, 15+, 15+, 16

Next time point t = 13

- Prob of making it to 10 months = 0.583
- Prob of making to 13 months given you have made it to 10 months = ?

3, 5, 7, 8+, 10, 13, 13+, 15+, 15+, 16

Next time point t = 13

- Prob of making it to 10 months = 0.583
- Prob of making to 13 months given you have made it to 10 months = (4/5)
- Prob of making to 13 months = 0.583*(4/5) = 0.467

3, 5, 7, 8+, 10, 13, 13+, 15+, 15+, 16

Next time point t = 15

- Prob of making it to 13 months = 0.467
- Prob of making to 15 months given you have made it to 13 months = (3/3)
- Prob of making to 15 months = 0.467*(3/3) = 0.467

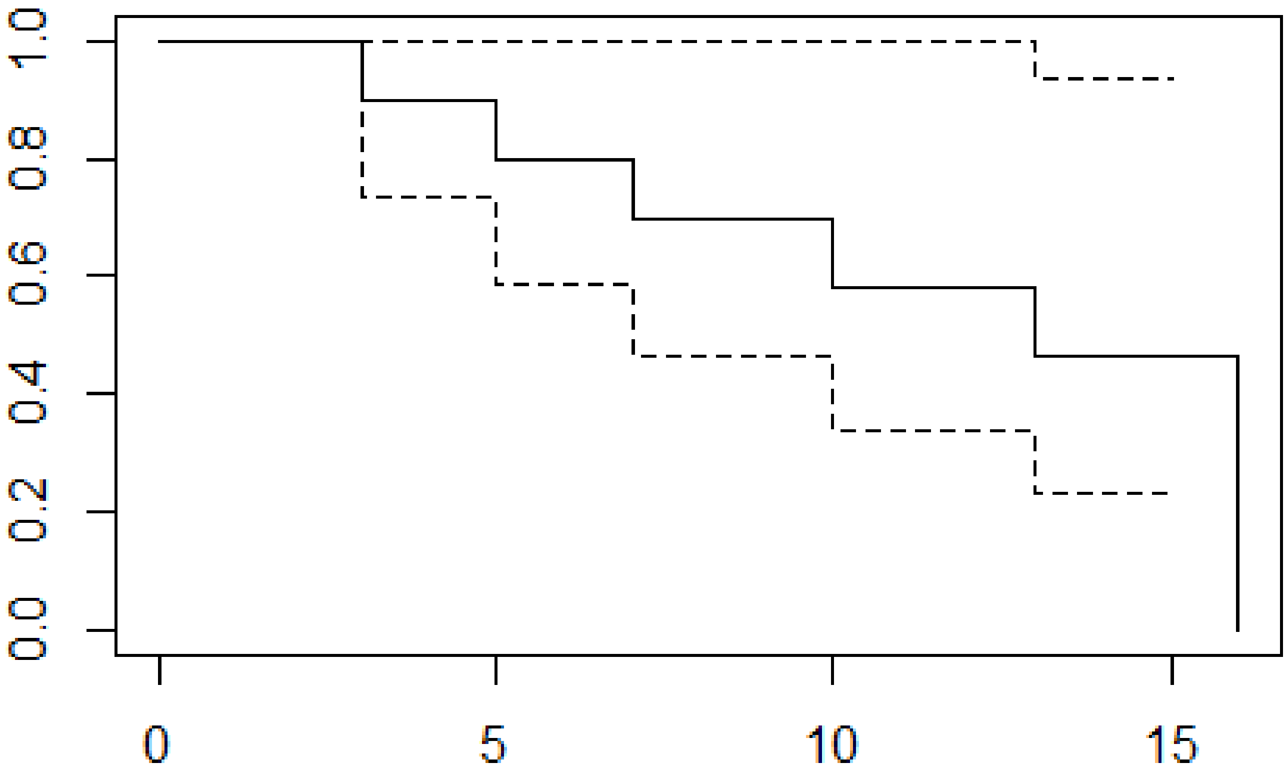
3, 5, 7, 8+, 10, 13, 13+, 15+, 15+, 16

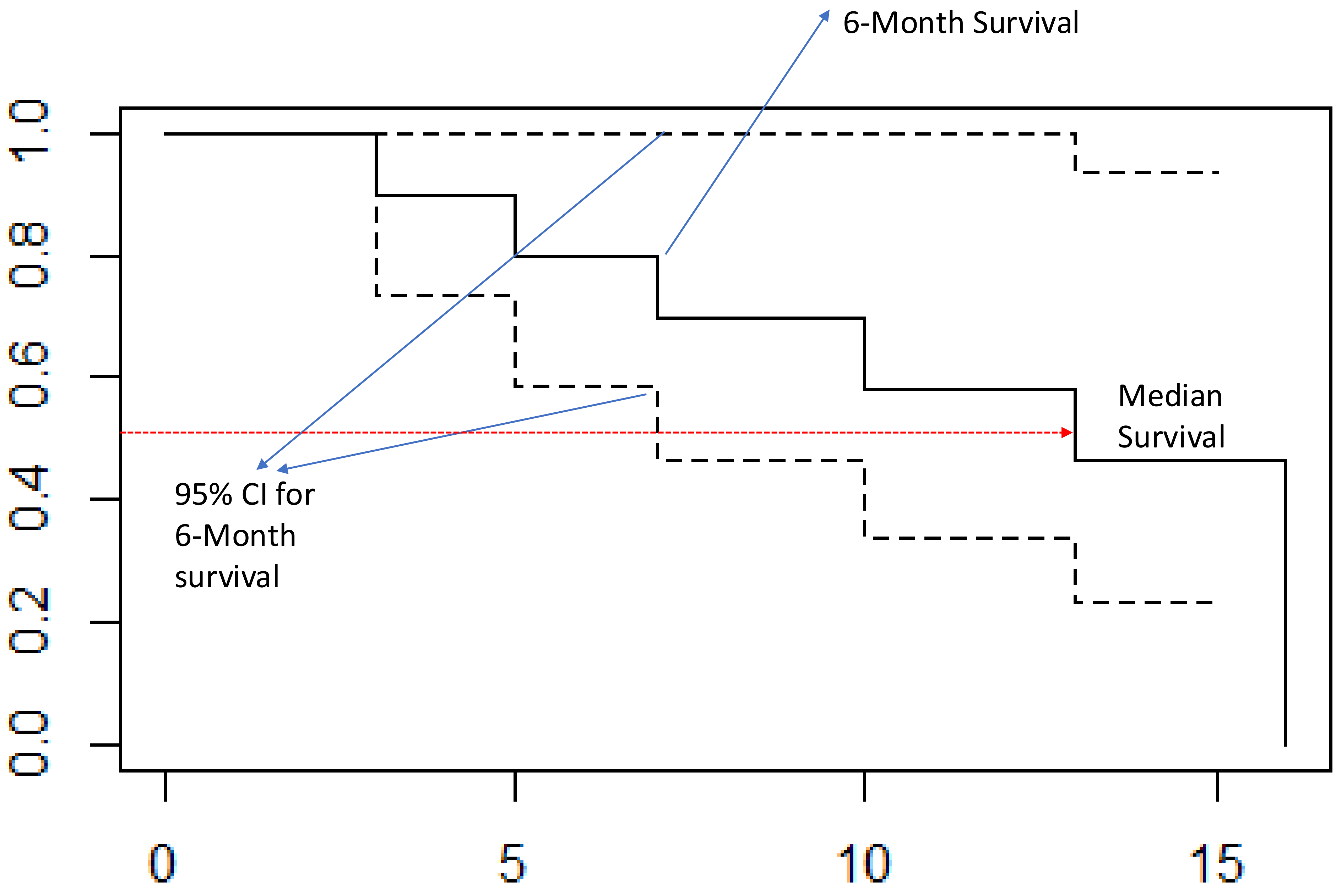
Next time point t = 16

- Prob of making it to 15 months = 0.467
- Prob of making to 16 months given you have made it to 16 months = (0/1)
- Prob of making to 16 months = 0.467*0 = 0

3, 5, 7, 8+, 10, 13, 13+, 15+, 15+, 16

Kaplan-Meier Curve !!!




6-Month Survival
1.0
0.8
0.6
0.4
0.2
0.0
Median
Survival
95% CI for
6-Month
survival
0
5
10
15

Characteristics of KM Curves

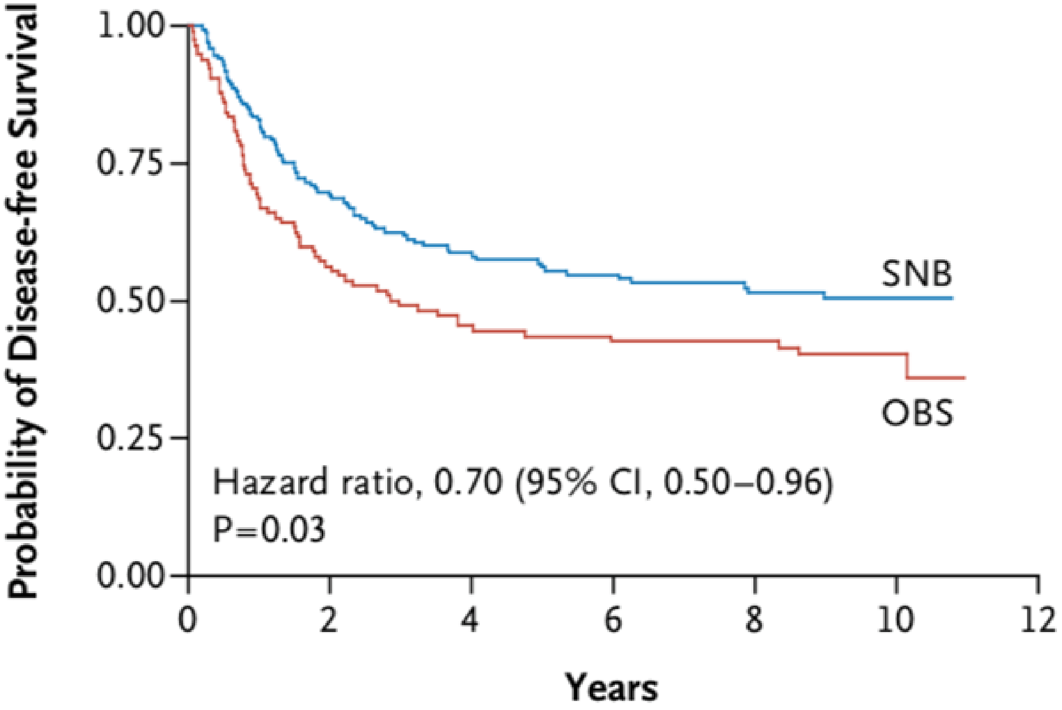
- Full information about survival
  - All probabilities
  - All quantiles, quartiles etc
- Jumps down at event times
- Stays the same at censoring times
  - This is why censoring times are indicated by tick marks
- Step sizes get larger and larger
  - Because the denominator gets smaller and smaller

Key Concept

- Risk set, or the number at risk changes at each time point
- This is true even when there is no censoring; those who die move out of the risk set
- But when censoring is present, those who are censored also move out of the risk set at the time of censoring
- The idea of risk set underlies all survival analysis (Cox regression, competing risks, truncation etc)

D **Disease-free Survival, Thick Melanomas**

	No. of Events/ Total No.	Rate (%) Yr 5	Rate (%) Yr 10
OBS	68/117	43.7±4.7	40.5±4.7
SNB	80/173	56.2±3.9	50.7±4.0



No. at Risk

OBS	117	63	49	44	40	20	0
SNB	173	116	92	75	58	36	0

Log-Rank Test

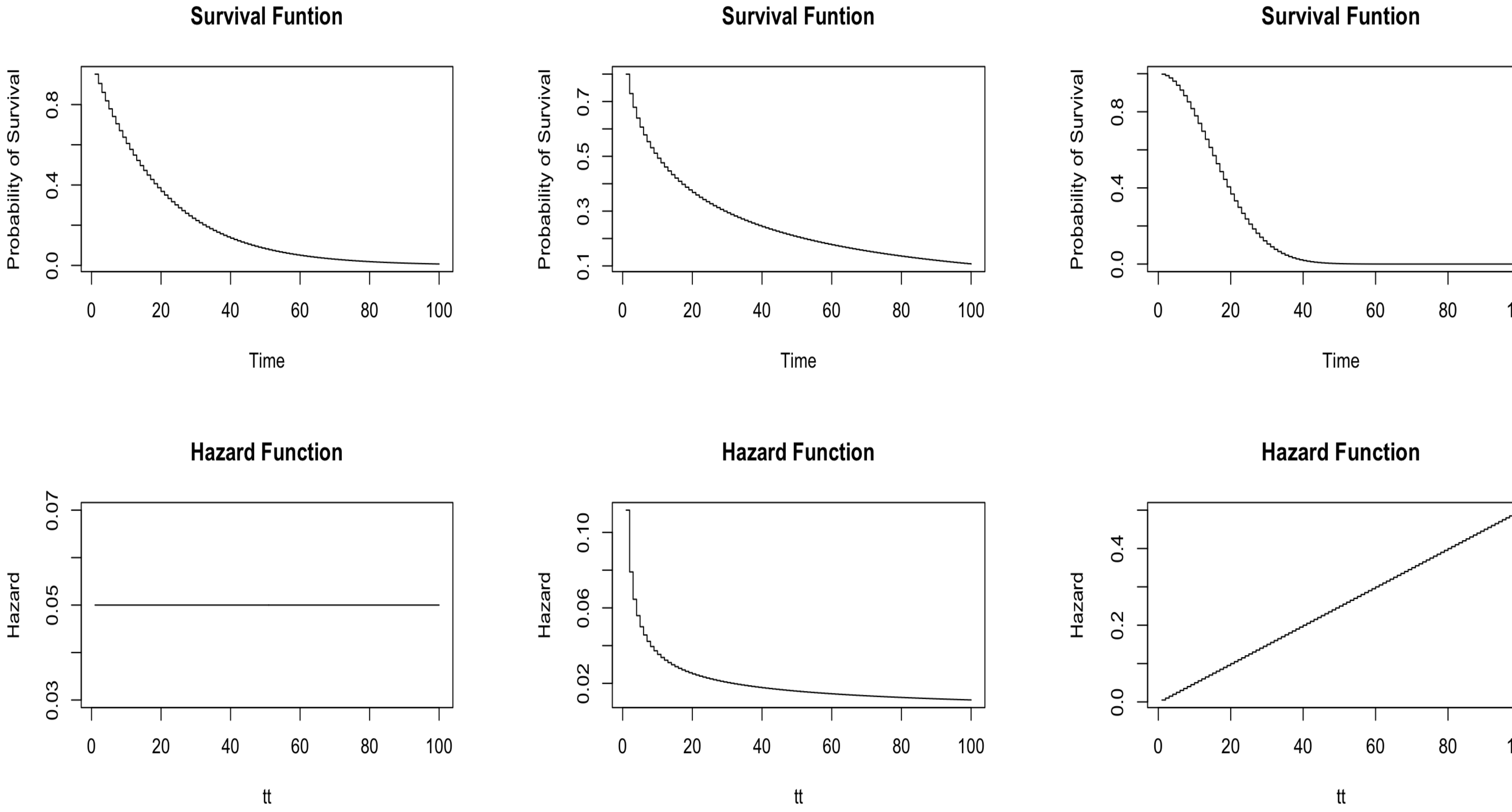
- Standard method of comparing survival curves
- At each time point one of the curves (or both) jumps, calculate the expected number of events at that time point if there is no difference between the curves
  - Why do we do this calculation assuming there is no difference?
- Compare the expected number with the observed (similar to the chi-square test)
- Add up the difference between observed and expected over all time points

Hazard Function

- Remember the survival curve, which represents one minus the cdf of the survival times?
- Call it S(t)
- Hazard function, h(t), is the derivative of the logarithm of S(t)
- $h(t) = (d/dt) \log(S(t))$

Hazard Function: Interpretation

- Probability of dying between t and t+d, given one has survived up to t, divided by d, where d is a very, very small time interval
- Hazard is NOT probability
- It is probability scaled by a small length of time
- Numerically, not useful at all by itself
- But the shape of the hazard function is quite informative

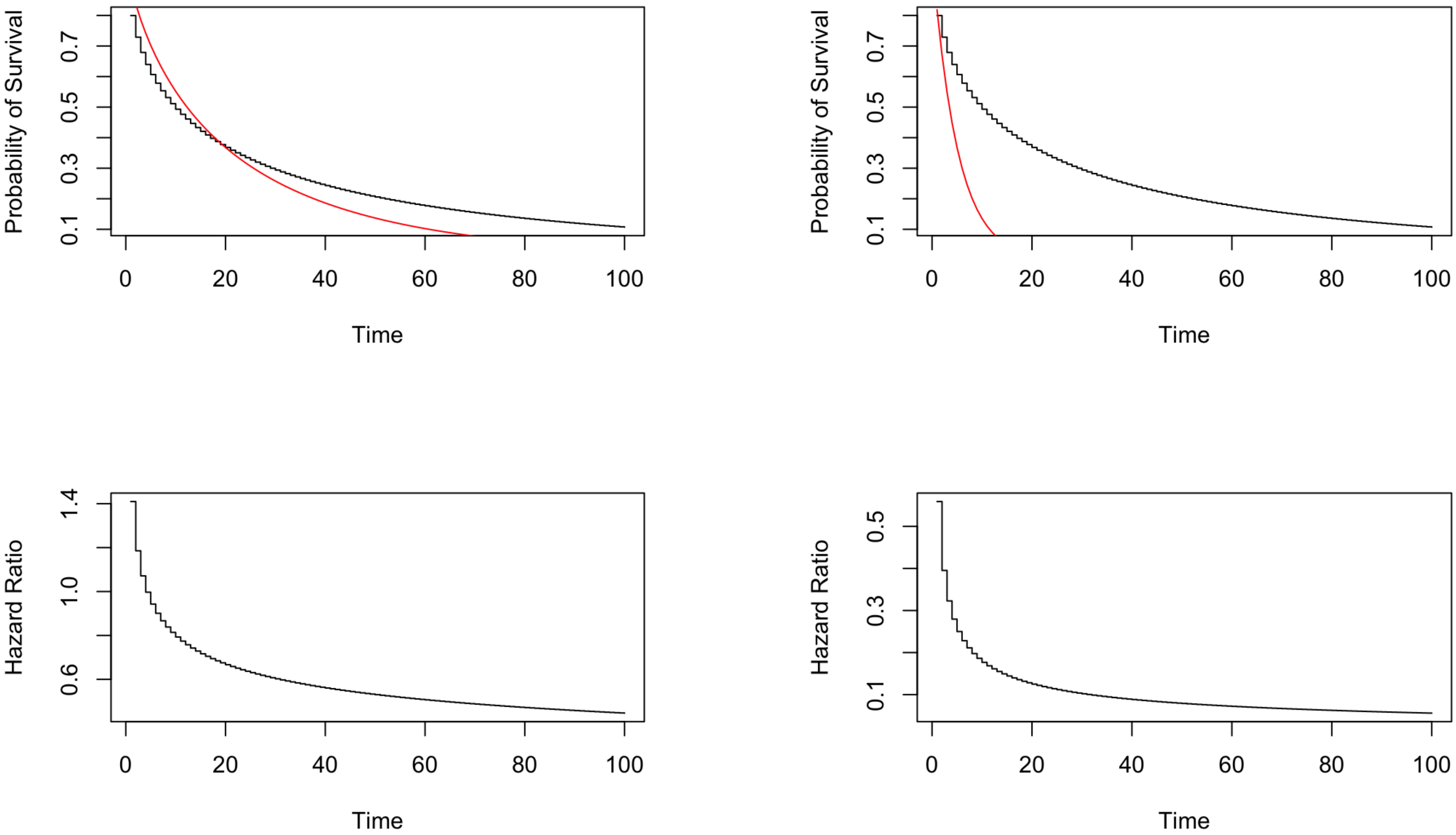

Survival Funtion
Probability of Survival
0.0
0.4
0.8
0
20
40
60
80
100
Time
Survival Funtion
Probability of Survival
0.1
0.3
0.5
0.7
0
20
40
60
80
100
Time
Survival Funtion
Probability of Survival
0.0
0.4
0.8
0
20
40
60
80
Time
Hazard Function
Hazard
0.03
0.05
0.07
0
20
40
60
80
100
tt
Hazard Function
Hazard
0.02
0.06
0.10
0
20
40
60
80
100
tt
Hazard Function
Hazard
0.0
0.2
0.4
0
20
40
60
80
tt

Three Examples

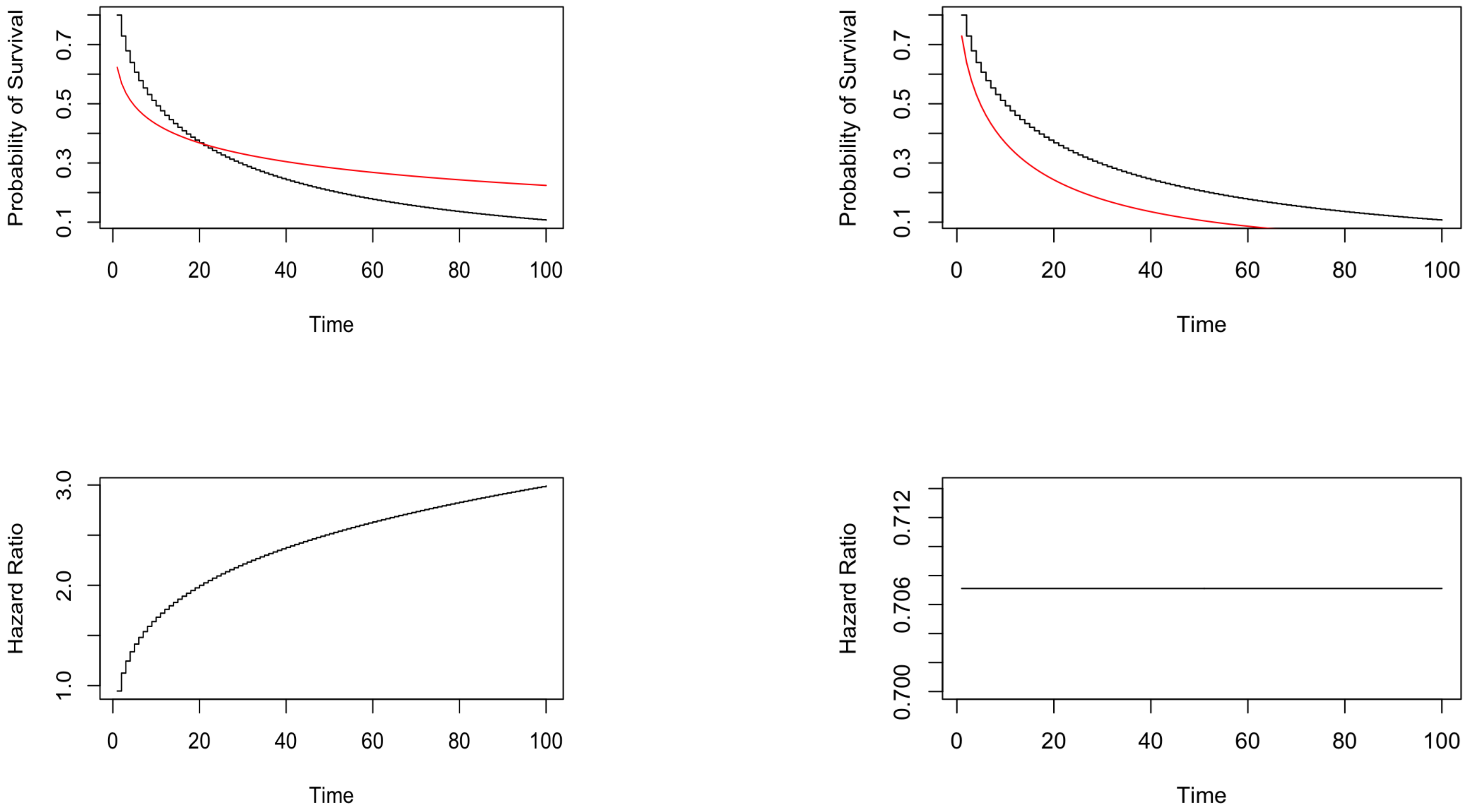
- Increasing hazard
  - Probability of death in population (over a long period of time)
- Decreasing hazard
  - Probability of cancer death or recurrence
- Constant hazard
  - ?

Hazard Ratio

- Hazard function itself is not used very much but ratio of two hazard functions is commonly used to “compare” survival functions



Probability of Survival
0.1
0.3
0.5
0.7
0
20
40
60
80
100
Time
Probability of Survival
0.1
0.3
0.5
0.7
0
20
40
60
80
100
Time
Hazard Ratio
0.6
1.0
1.4
0
20
40
60
80
100
Time
Hazard Ratio
0.1
0.3
0.5
0
20
40
60
80
100
Time


Probability of Survival
0.1
0.3
0.5
0.7
0
20
40
60
80
100
Time
Probability of Survival
0.1
0.3
0.5
0.7
0
20
40
60
80
100
Time
Hazard Ratio
1.0
2.0
3.0
0
20
40
60
80
100
Time
Hazard Ratio
0.700
0.706
0.712
0
20
40
60
80
100
Time

Hazard Ratio

- Hazard function itself is not used very much but ratio of two hazard functions is commonly used to “compare” survival functions
- Constant hazard ratio → Proportional hazards (PH)
- Crossing survival curves → Not PH
- Converse is not true, one can have PH without survival curves crossing
- Eyeballing the survival curves to figure out hazard ratios is futile

Restricted Mean Survival Time (RMST)

- Fact: Area under the survival curve is the mean survival time
- Why not used more often (or at all)?
- Unless survival curve drops to 0, you cannot estimate the mean
- Idea: choose a time point on your survival curve that is both clinically relevant and also “well-estimated” (i.e. still a reasonable number at risk)
- Calculate the area under the curve up to that time point
- → RMST

RMST Difference

- If the area under one curve is the mean
- Then the area between the curves is mean difference
- The are between the curves up to a time point is RMST Difference
- An alternative to HR to summarize the relative position of the curves
- Does not make assumptions
- But requires a “subjective” choice of time point
  - Although less influenced by this choice than, say, reporting S(t) at a chosen time point

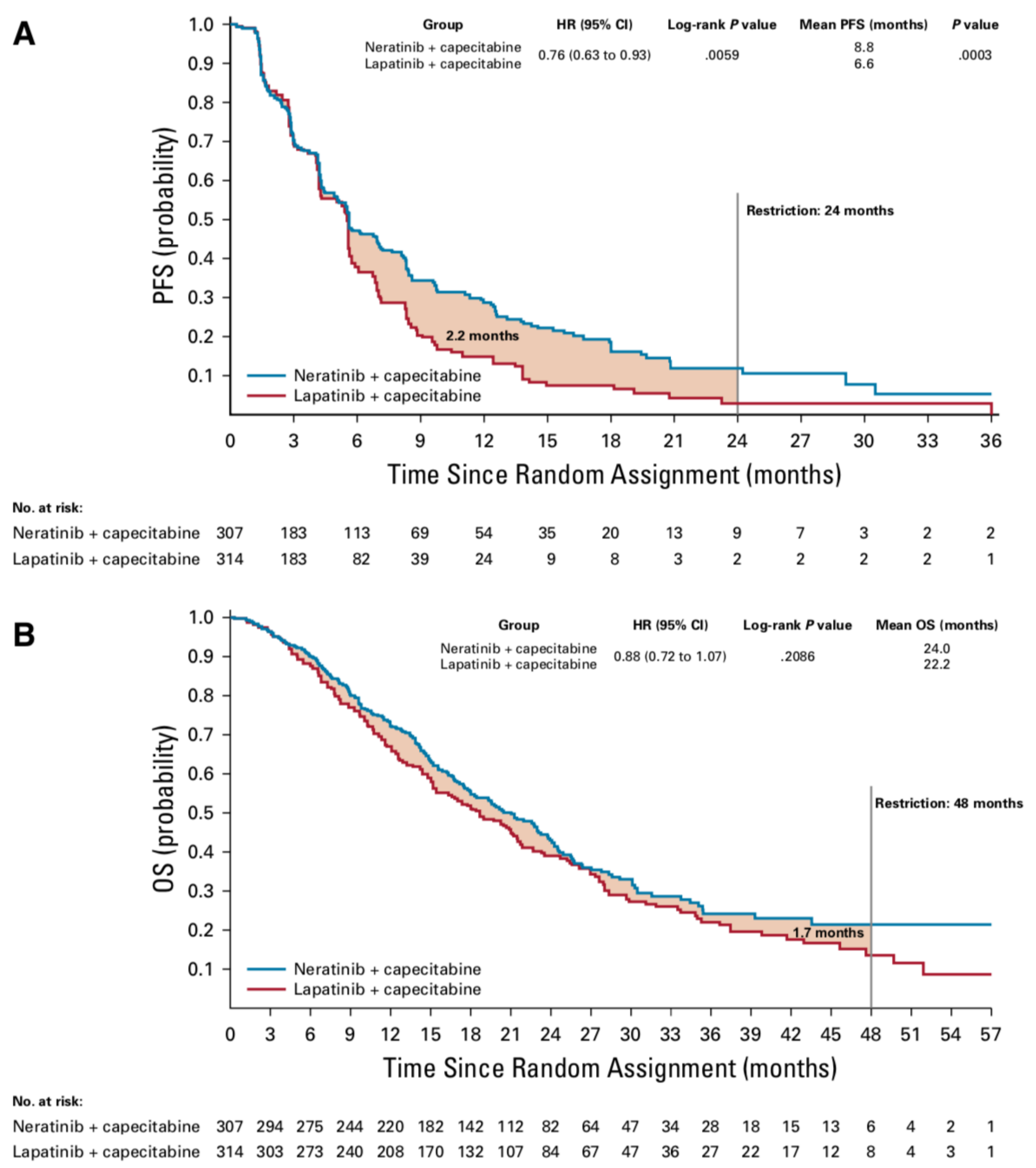
RMST Difference is most useful

- When PH does not hold
- Caveats:
- How can you tell PH does not hold?
  - Very difficult (unless curves cross)
  - The statistical tests developed for this are not powerful
- Need to be careful if curves cross
  - You will need to keep track of which curve is on top where and assign “a sign” to the area

original reports

Neratinib Plus Capecitabine Versus Lapatinib Plus Capecitabine in HER2-Positive Metastatic Breast Cancer Previously Treated With ≥ 2 HER2-Directed Regimens: Phase III NALA Trial

Cristina Saura, MD[1]; Mafalda Oliveira, MD, PhD[1]; Yin-Hsun Feng, MD, PhD[2]; Ming-Shen Dai, MD, PhD[2]; Shang-Wen Chen, MD[2]; Sara A. Hurvitz, MD[3]; Sung-Bae Kim, MD, PhD[4]; Beverly Moy, MD, PhD[5]; Suzette Delaloge, MD, MSc[6]; William Gradishar, MD[7]; Norikazu Masuda, MD, PhD[8]; Marketa Palacova, MD[9]; Maureen E. Trudeau, MD[10]; Johanna Mattson, MD, PhD[11]; Yoon Sim Yap, MBBS[12]; Ming-Feng Hou, MD[13]; Michelino De Laurentiis, MD, PhD[14]; Yu-Min Yeh, MD[15]; Hong-Tai Chang, MD[16]; Thomas Yau, MBBS, MD[17]; Hans Wildiers, MD, PhD[18,19]; Barbara Haley, MD[20]; Daniele Fagnani, MD[21]; Yen-Shen Lu, MD, PhD[22]; John Crown, MBBCh, MD[23]; Johnson Lin, MD[24]; Masato Takahashi, MD, PhD[25]; Toshimi Takano, MD[26]; Miki Yamaguchi, MD, PhD[27]; Takaaki Fujii, MD, PhD[28]; Bin Yao, MS[29]; Judith Bebchuk, ScD[29]; Kiana Keyvanjah, PharmD[29]; Richard Bryce, MBChB[29]; and Adam Brufsky, MD, PhD[30]; for the NALA Investigators


A
Group
HR (95% CI)
Log-rank P value
Mean PFS (months)
P value
Neratinib + capecitabine
Lapatinib + capecitabine
0.76 (0.63 to 0.93)
.0059
8.8
6.6
.0003
Restriction: 24 months
2.2 months
PFS (probability)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Neratinib + capecitabine
Lapatinib + capecitabine
0 3 6 9 12 15 18 21 24 27 30 33 36
Time Since Random Assignment (months)
No. at risk:
Neratinib + capecitabine 307 183 113 69 54 35 20 13 9 7 3 2 2
Lapatinib + capecitabine 314 183 82 39 24 9 8 3 2 2 2 2 1
B
Group
HR (95% CI)
Log-rank P value
Mean OS (months)
Neratinib + capecitabine
Lapatinib + capecitabine
0.88 (0.72 to 1.07)
.2086
24.0
22.2
Restriction: 48 months
1.7 months
OS (probability)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Neratinib + capecitabine
Lapatinib + capecitabine
0 3 6 9 12 15 18 21 24 27 30 33 36 39 42 45 48 51 54 57
Time Since Random Assignment (months)
No. at risk:
Neratinib + capecitabine 307 294 275 244 220 182 142 112 82 64 47 34 28 18 15 13 6 4 2 1
Lapatinib + capecitabine 314 303 273 240 208 170 132 107 84 67 47 36 27 22 17 12 8 4 3 1