ATLAS
ASIAN CLINICAL TRIALS NETWORK FOR CANCERS PROJECT

National Cancer Center Japan

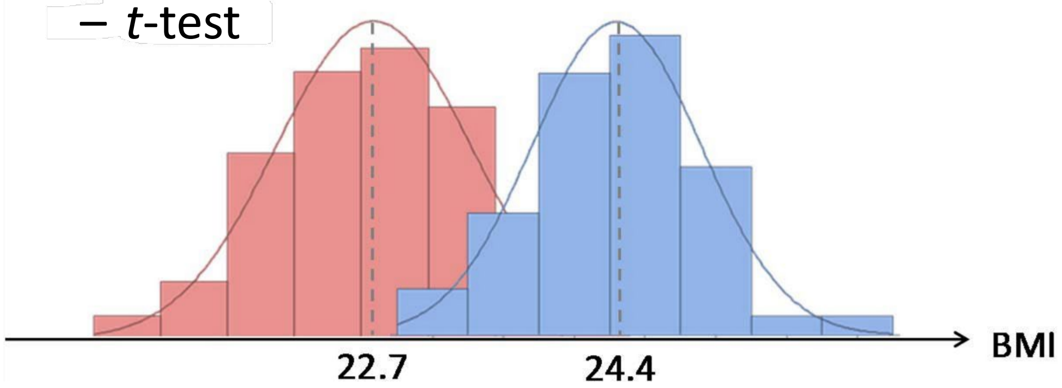
Survival Analysis

National Cancer Center, Japan
Biostatistics Division
Junki Mizusawa

JCOG
Japan Clinical Oncology Group

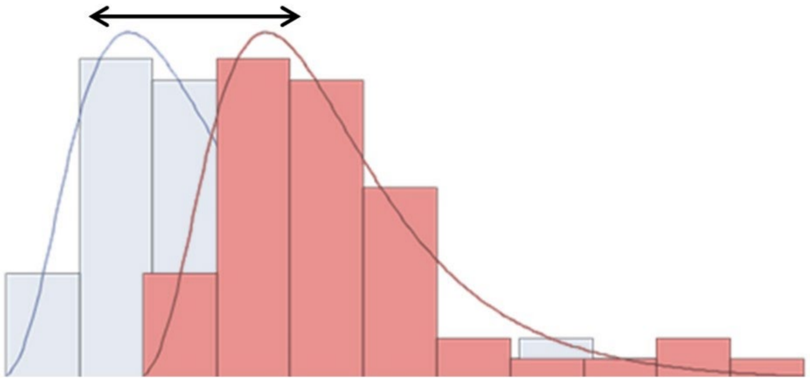
Review of the Previous Lecture: Analysis of Continuous Data

- Step 1: Confirm distribution with histogram
- Step 2: Group comparison
  - *t*-test



$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{V\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$

  - Wilcoxon rank sum test



	BMI (kg/m^2)				Mean
Group A	22.0 [2]	28.3 [8]	19.4 [1]	22.3 [3]	23.0
Group B	23.3 [4]	25.1 [7]	24.6 [6]	23.5 [5]	24.1

Review of the Previous Lecture: Analysis of Binary Data

- Step 1: Create a (2 × 2) contingency table and tabulate frequencies

		Onset of colorectal cancer		
		Yes	No	Total
Aspirin	Yes	129	4932	5061
	No	87	2440	2527
	Total	216	7372	7588

- Step 2: Group comparison
  - Fisher's exact test
  - One-sided P value = 0.0053 + … + 1.2 × 10−106 = 0.0174

216	4845	5061
0	2527	2527
216	7372	7588

$Pr = 2.4 \times 10^{-37}$

215	4846	5061
1	2526	2527
216	7372	7588

$Pr = 1.3 \times 10^{-35}$

214	4847	5061
2	2525	2527
216	7372	7588

$Pr = 5.0 \times 10^{-34}$

・・・

129	4932	5061
87	2440	2527
216	7372	7588

Pr = 0.0053

・・・

0	5061	5061
216	2311	2527
216	7372	7588

$Pr = 1.2 \times 10^{-106}$

The same contingency table as the observed data

Review of the Previous Lecture: Analysis of Binary Data

- Step 1: Create a (2 × 2) contingency table and tabulate frequencies

		Onset of colorectal cancer		
		Yes	No	Total
Aspirin	Yes	129	4932	5061
	No	87	2440	2527
	Total	216	7372	7588

- Step 2: Group comparison
  - Chi-square test

Observed values	
129 (2.5%)	4932 (97.5%)
87 (3.4%)	2440 (96.6%)

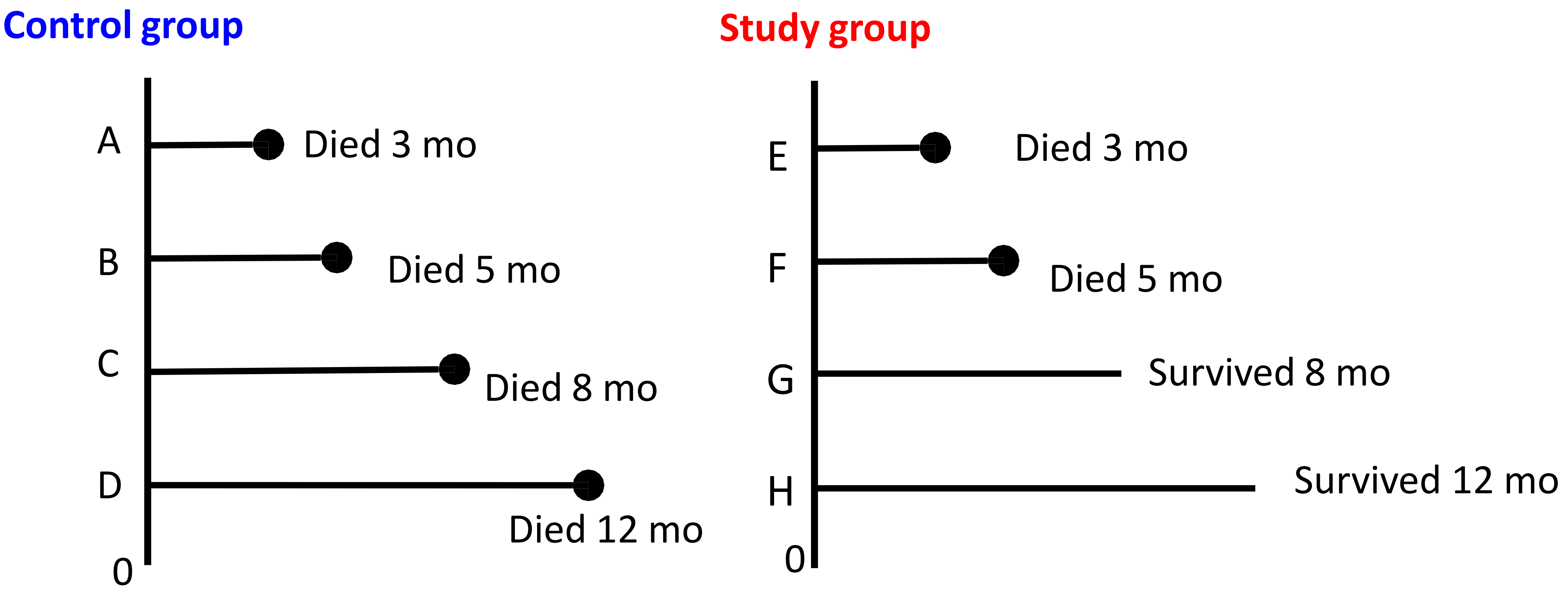
Expected values	
144 (2.8%)	4917 (97.2%)
72 (2.8%)	2455 (97.2%)

- Test based on the difference between observed and expected frequencies
  - The greater the difference between the observed and expected frequencies, the rarer the result observed under the null hypothesis.
- Statistic $\frac{(1239-144)^2}{144}+\frac{(4932-4917)^2}{4917}+\frac{(87-72)^2}{72}+\frac{(2440-2455)^2}{2455}$
- Approximates a chi-square distribution with one degree of freedom
  - The higher the expected frequency, the better the approximation.
- P = 0.0273 (Significant at the two-sided 5% level!)

Outcome Types and Statistical Methods

	Continuous Variable	Binary (0/1)	Survival Time
Outcome examples	BMI, blood pressure, laboratory values	Response rate	Overall survival, progression-free survival
Data summary	Histogram	Contingency table	?
Group comparison (test)	*t*-test, Wilcoxon test	Chi-square test, Fisher's exact test	?
Model fitting	Multiple regression analysis	Logistic regression	?

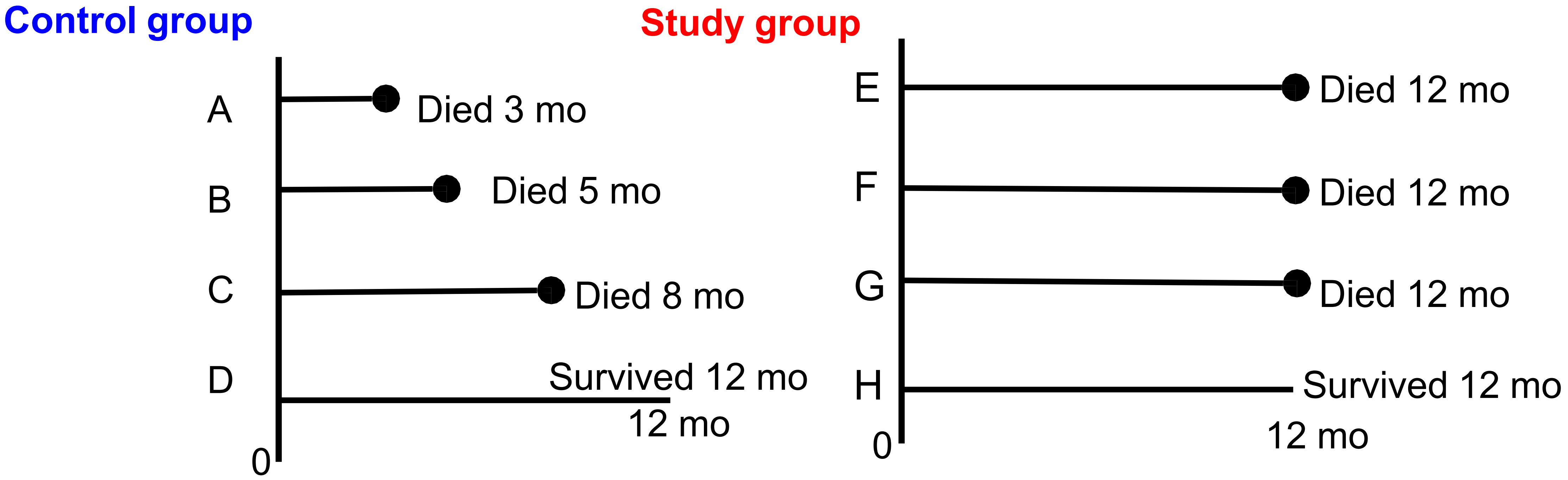
Question 1: Are Mean and *t*-test Sufficient?



- Mean of control group: (3 + 5 + 8 + 12)/4 = 7 mo
- Mean of study group: (3 + 5 + 8 + 12)/4 = 7 mo??

The survival in the study group is expected to be longer

Question 2: Is It Correct to Compare (Chi-square Test and Fisher's Exact Test) the Survival Rates at the 12-Month Time Point?

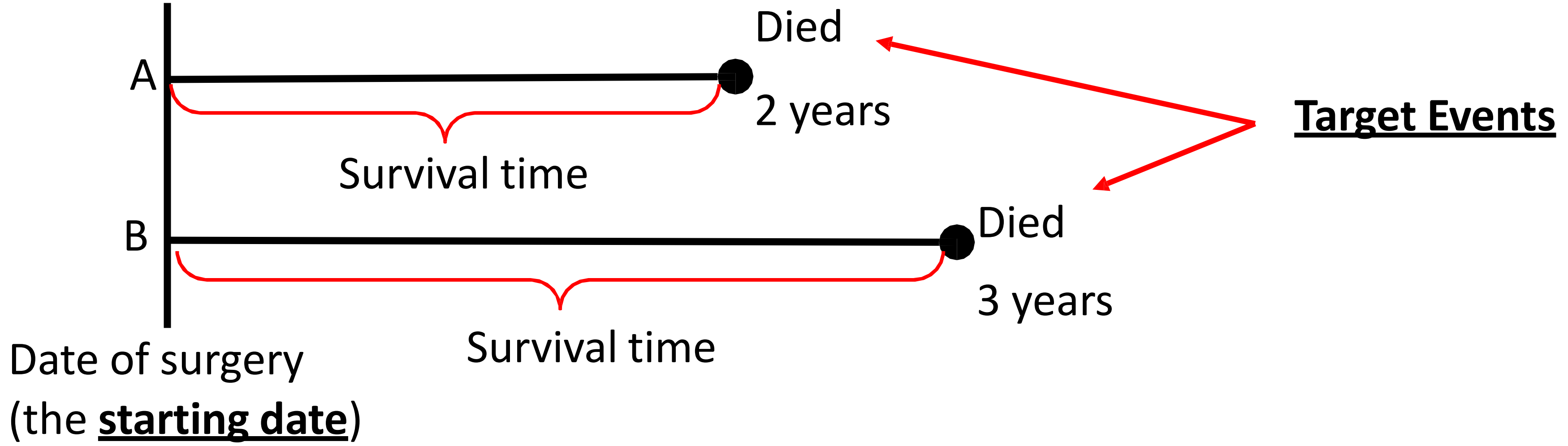


- Survival rate of control group: 1/4 = 25%
- Survival rate of study group: 1/4 = 25% ??

The survival in the study group is expected to be longer

What is Survival Analysis?

- What is survival time?
  - The **time** from a specific time point (the **starting date**) until the target **event** occurs.
    - From the date of enrollment in the clinical study to the date of death
    - From the date of surgery to the date of recurrence



- The analytical method used when interested in survival time is called **survival analysis**.

Survival Analysis for Clinical Study

Lung cancer internal medicine group JCOG0301

Unresectable stage III NSCLC
≥71 years

↓

Random assignment

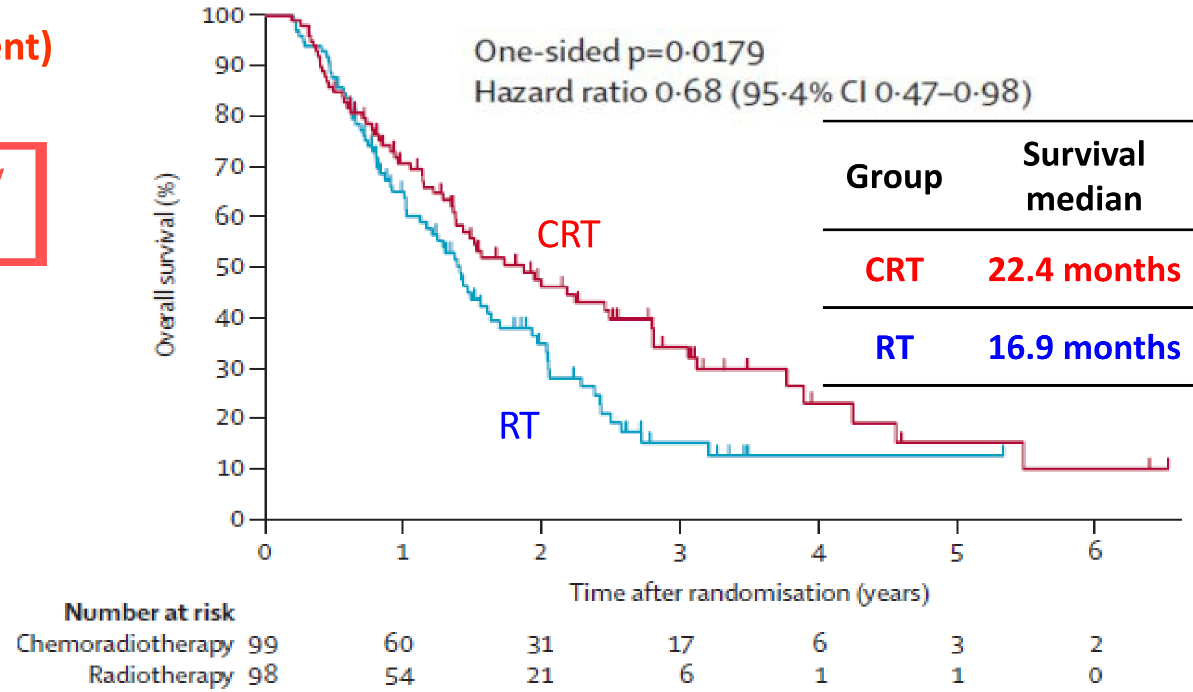
(Standard of care) 100 cases → Radiation alone (RT)

(New treatment) 100 cases → Radiation + chemotherapy (CRT)

Atagi *et al.* (2012) Lancet Oncol. 13(7): 671-8.

Statistical analysis

The primary endpoint of overall survival was analysed by the stratified log-rank test for eligible patients with performance status (0 *vs* 1–2 [because of the small number of patients with performance status of 2]) and stage (IIIA *vs* IIIB) as stratification factors. Overall survival and PFS were estimated using the Kaplan-Meier method; hazard ratios (HRs) were calculated by Cox regression, and response rates were compared by the



Group	Survival median	2 years survival rate
CRT	22.4 months	46.3%
RT	16.9 months	35.1%

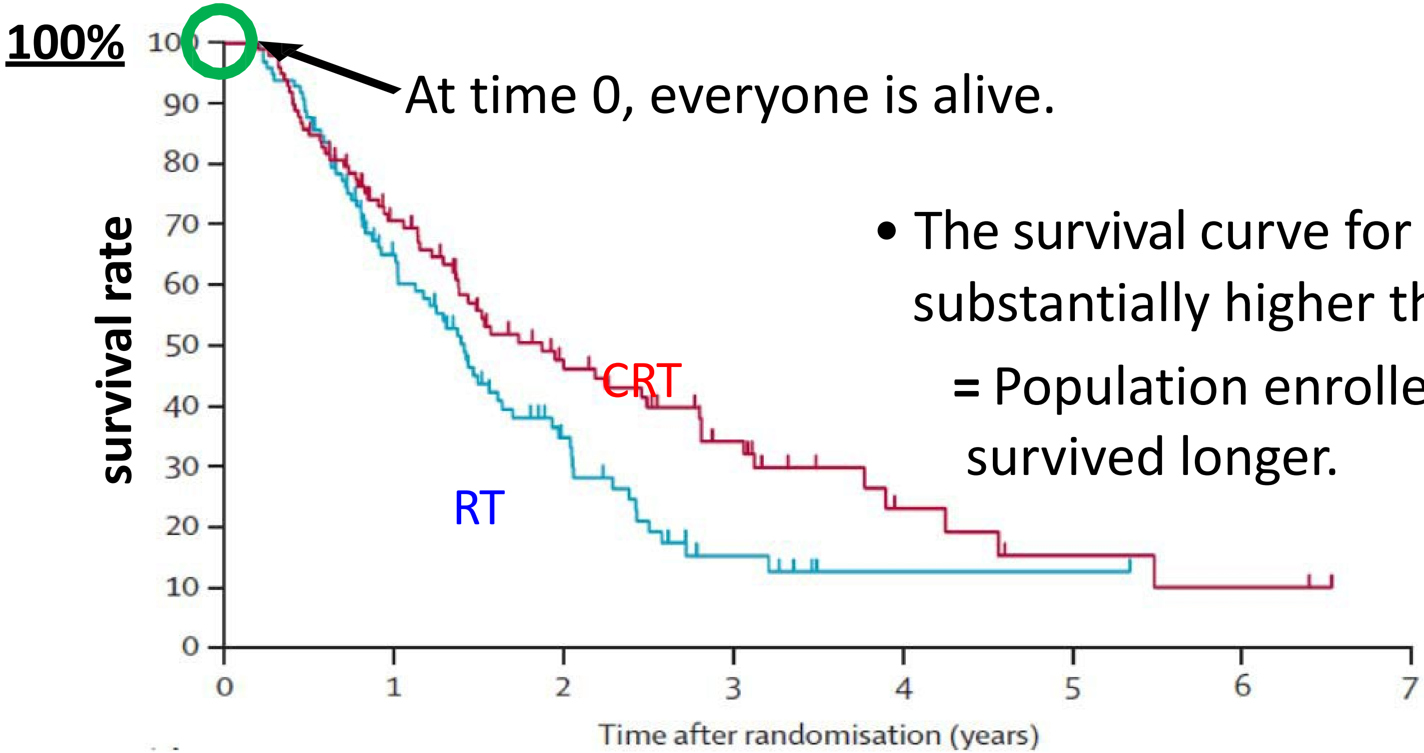
Learn the Three Essentials of Survival Analysis

- How to estimate and interpret the survival curve
  - **Kaplan–Meier method**
  - Number at risk
- Comparison of survival curves between groups
  - **Log-rank test**
- Estimation of treatment effect between groups of survival curves
  - **Cox regression** and proportional hazards
- Differences between rate, ratio, and proportion

Survival Curve

What Is a Survival Curve?

- The vertical axis represents the survival rate at each time point, the horizontal axis represents time, and the survival rate at each time point in the population is connected.
- When there is a death, the survival rate is reduced at that time.

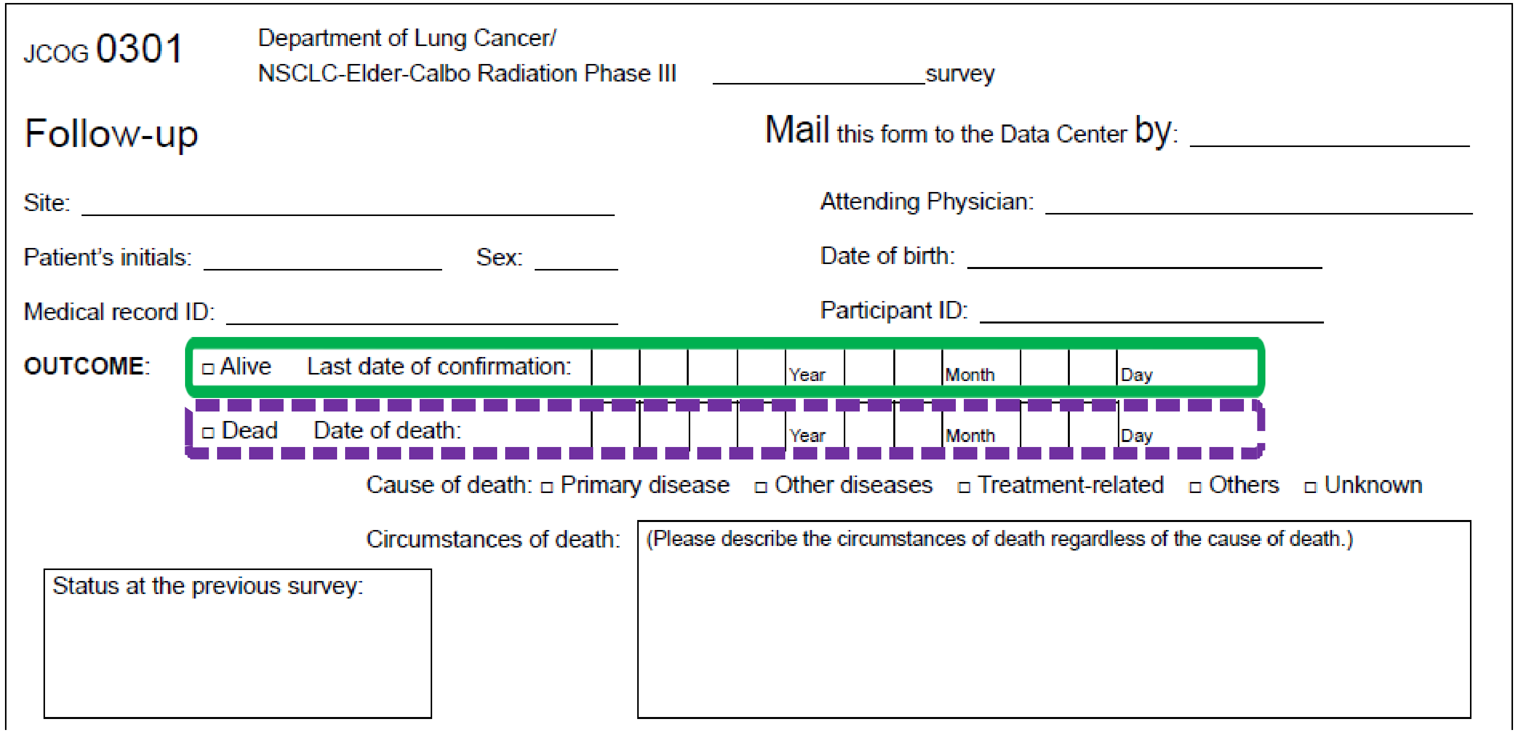


- The survival curve for the **CRT group** is substantially higher than that for the **RT group**.

  = Population enrolled in the **CRT group** survived longer.

Atagi *et al.* (2012) Lancet Oncol. 13(7): 671-8.

What Is a Survival Curve?



JCOG 0301 Department of Lung Cancer/ NSCLC-Elder-Calbo Radiation Phase III ______________survey

Follow-up

Mail this form to the Data Center by: ______________

Site: ______________ Attending Physician: ______________

Patient's initials: ______________ Sex: ______ Date of birth: ______________

Medical record ID: ______________ Participant ID: ______________

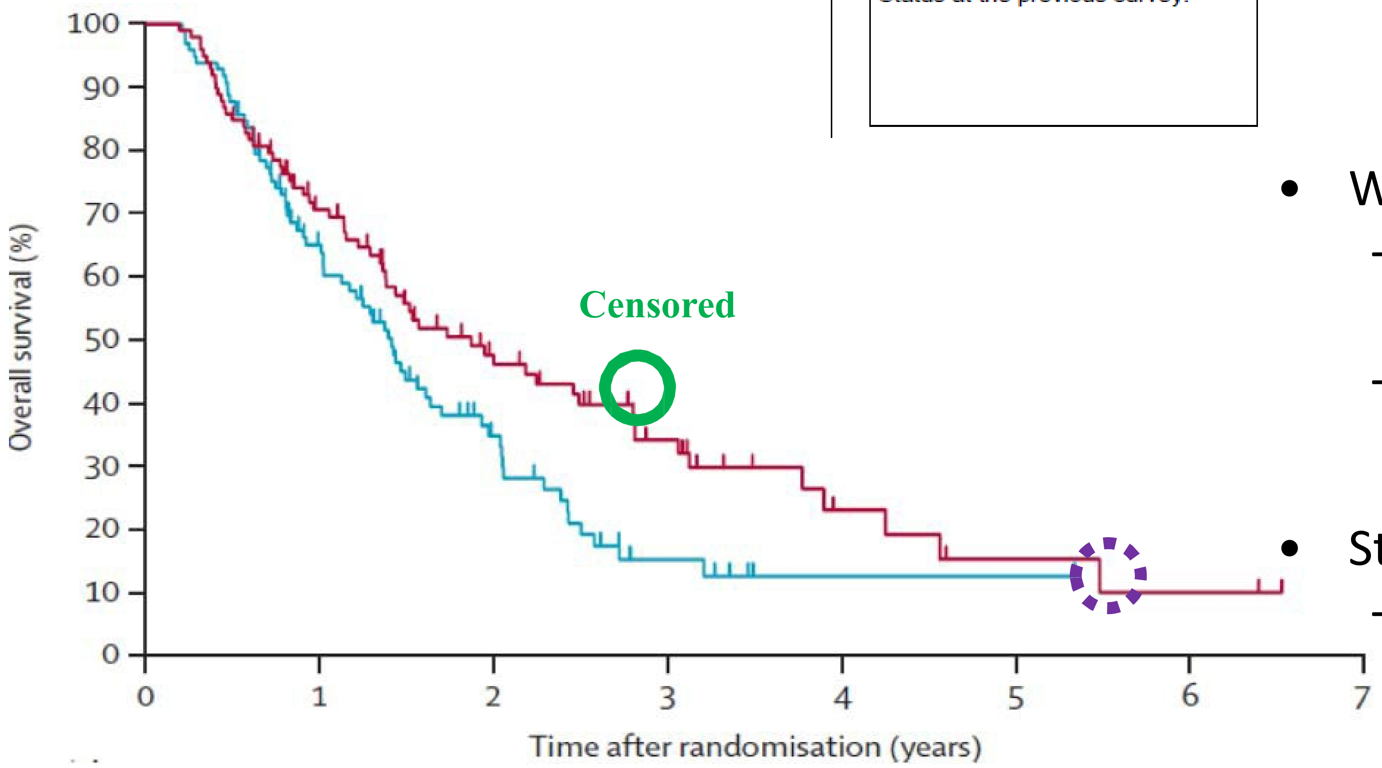
OUTCOME: ◻ Alive Last date of confirmation: Year Month Day

◻ Dead Date of death: Year Month Day

Cause of death: ◻ Primary disease ◻ Other diseases ◻ Treatment-related ◻ Others ◻ Unknown

Circumstances of death: (Please describe the circumstances of death regardless of the cause of death.)

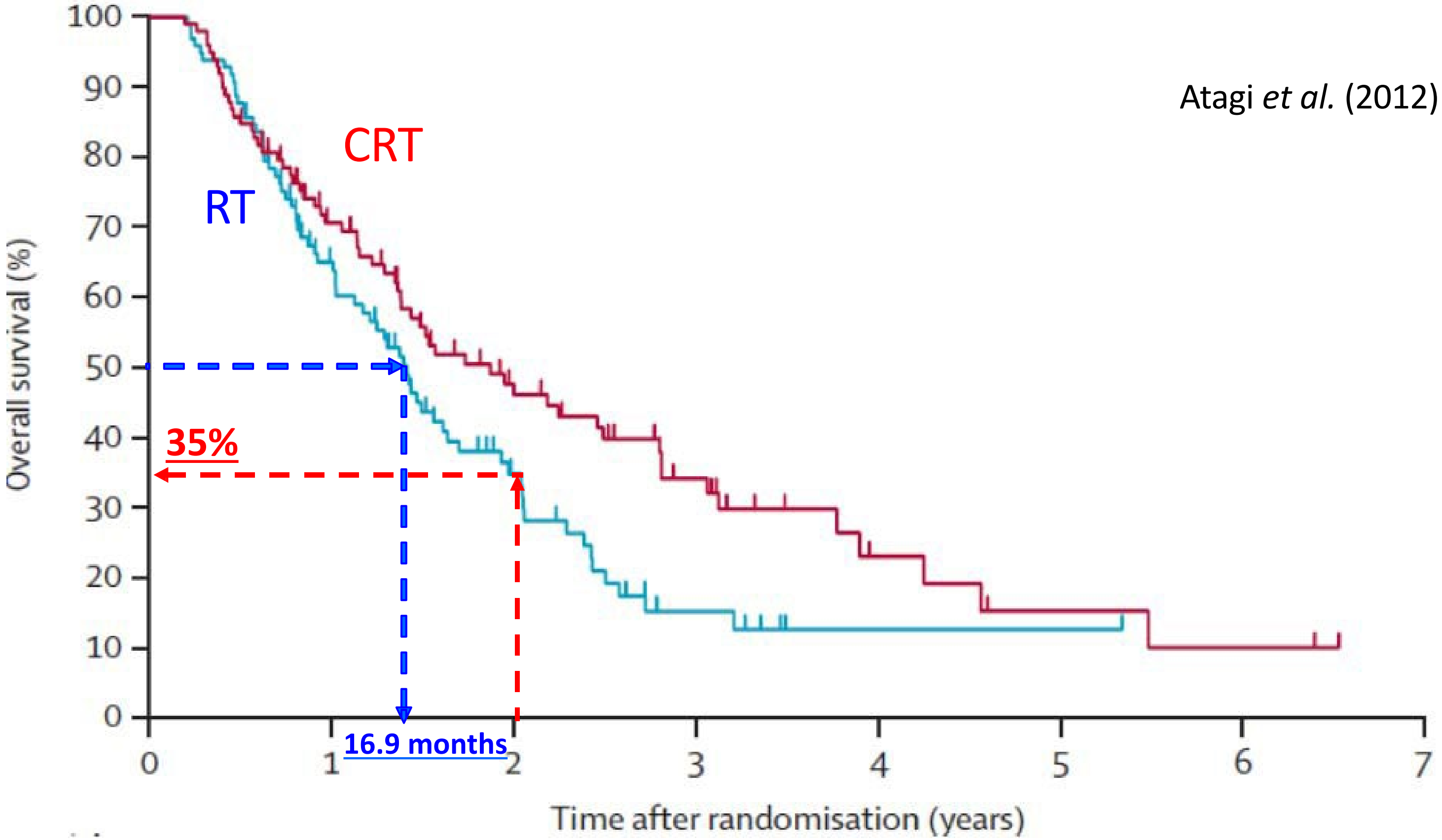
Status at the previous survey:



- Whiskers indicate censored cases
  - A patient who had no events up to that point but no further information is available.
  - Censored at the date of the last confirmation of survival
- Steps represent deaths.
  - Events at the date of death

Summary Values Obtained from Survival Curves

- Median survival time (MST), annual survival rate
  - The MST of the RT group was **16.9 months** and the 2-year survival rate was **35%**.

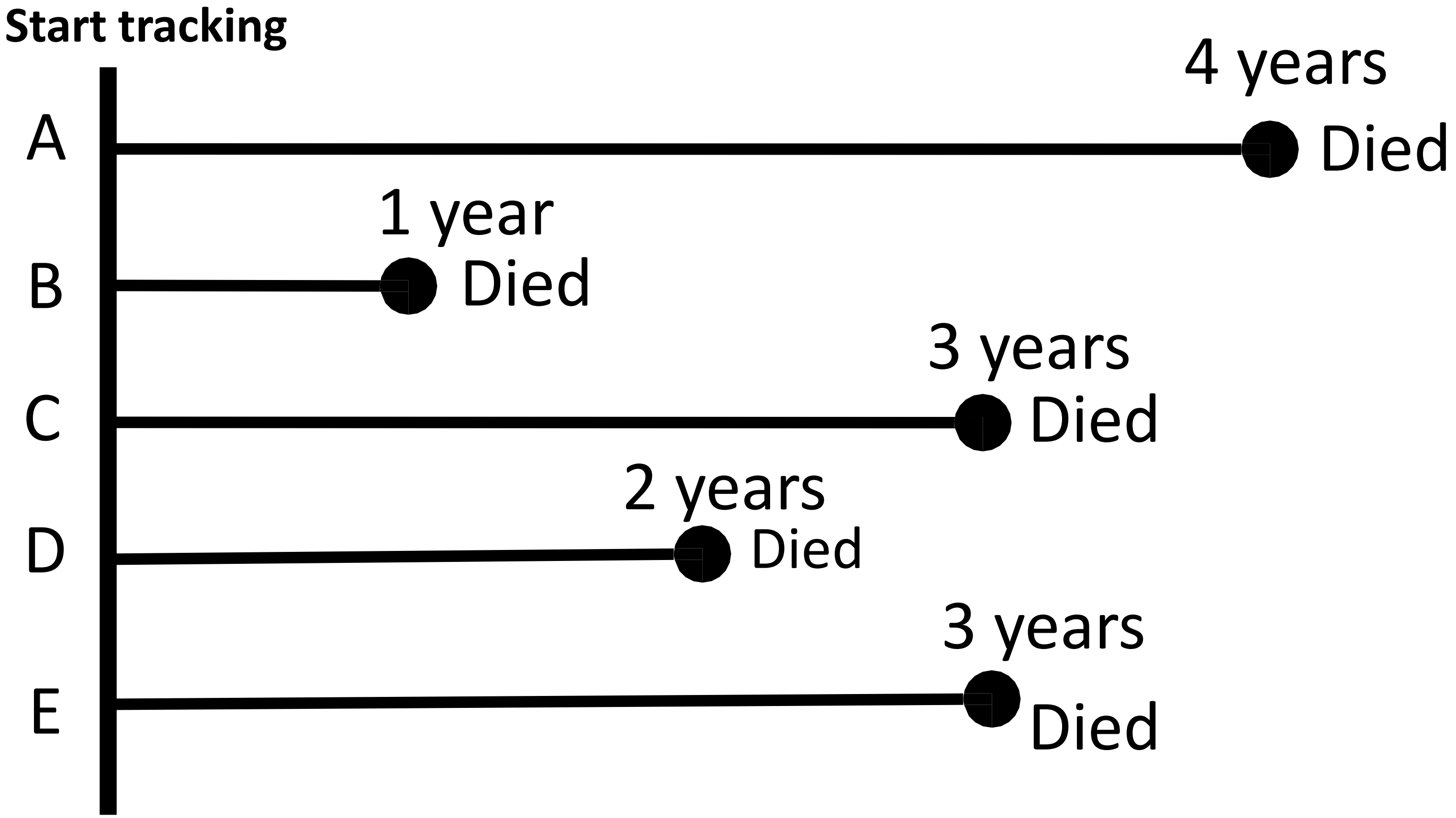


Atagi *et al.* (2012) Lancet Oncol. 13(7): 671-8.

Kaplan-Meier Method

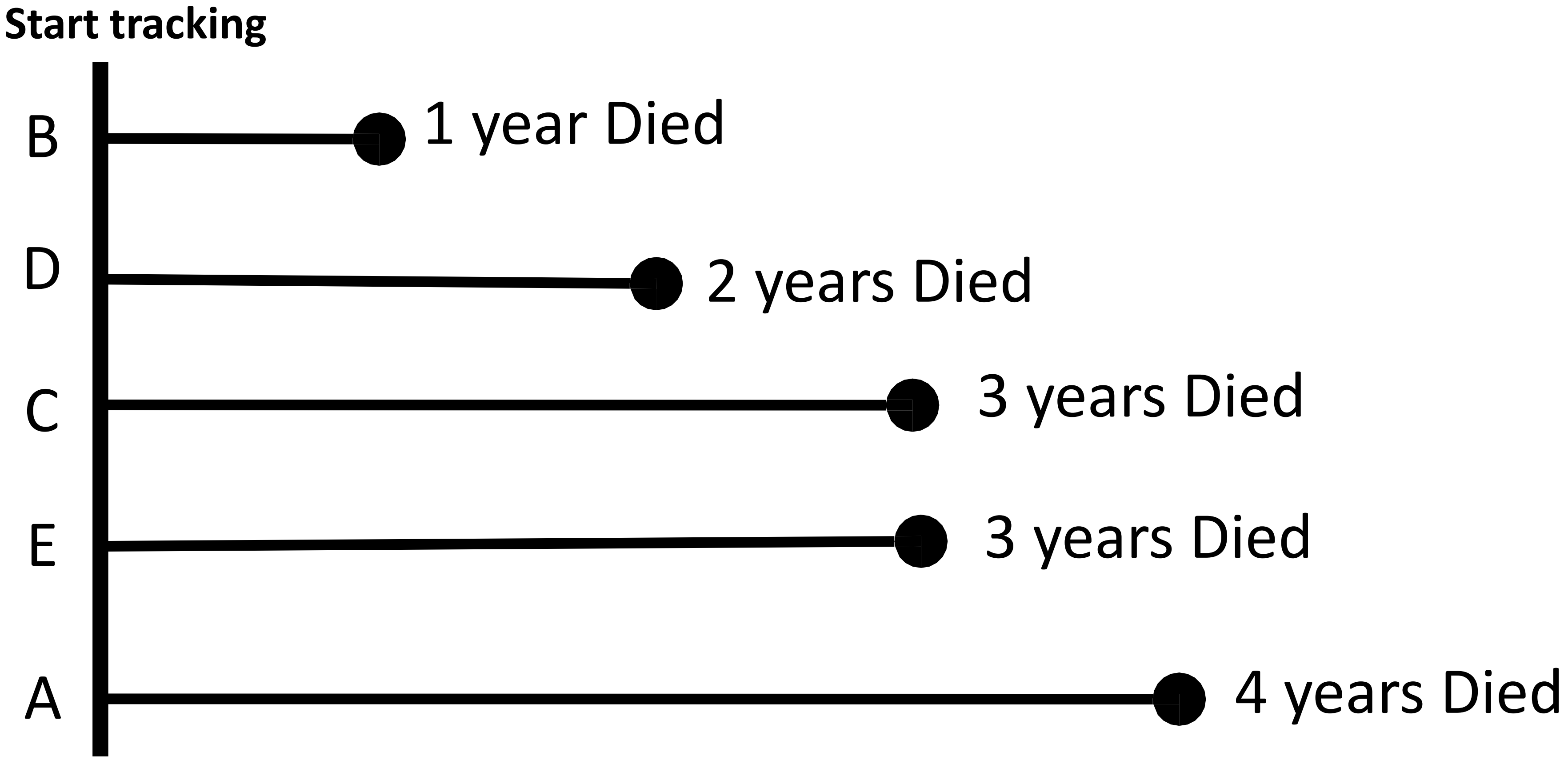
Example 1: When There Is No Censoring

- Estimation of the survival curve by the Kaplan–Meier method



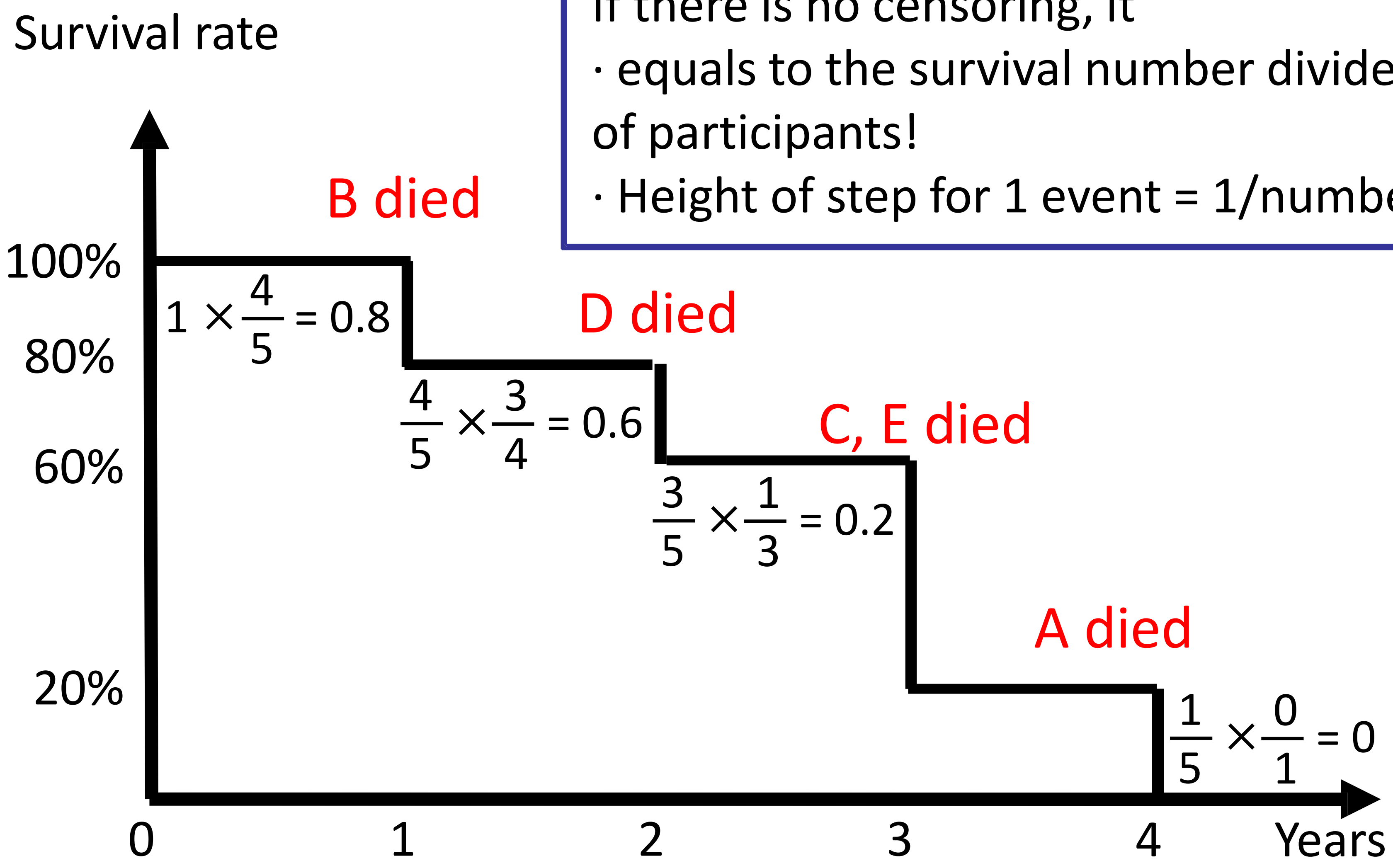
Rearrange in the Order of the Shortest Survival to Longest

- In short order, B → D → C , E → A



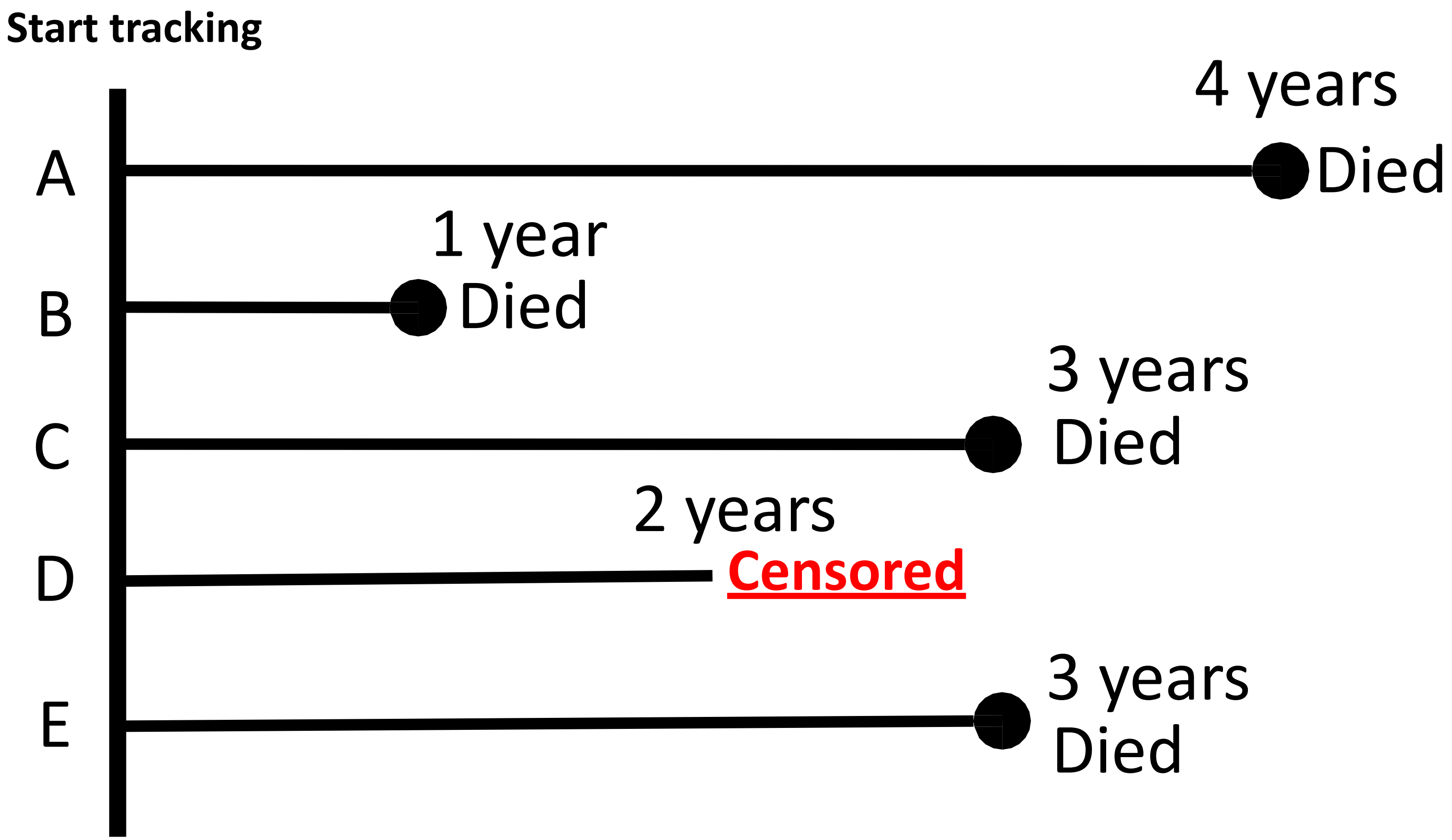
Kaplan–Meier Curve (KM Curve)

If there is no censoring, it
· equals to the survival number divided by the number of participants!
· Height of step for 1 event = 1/number of participants



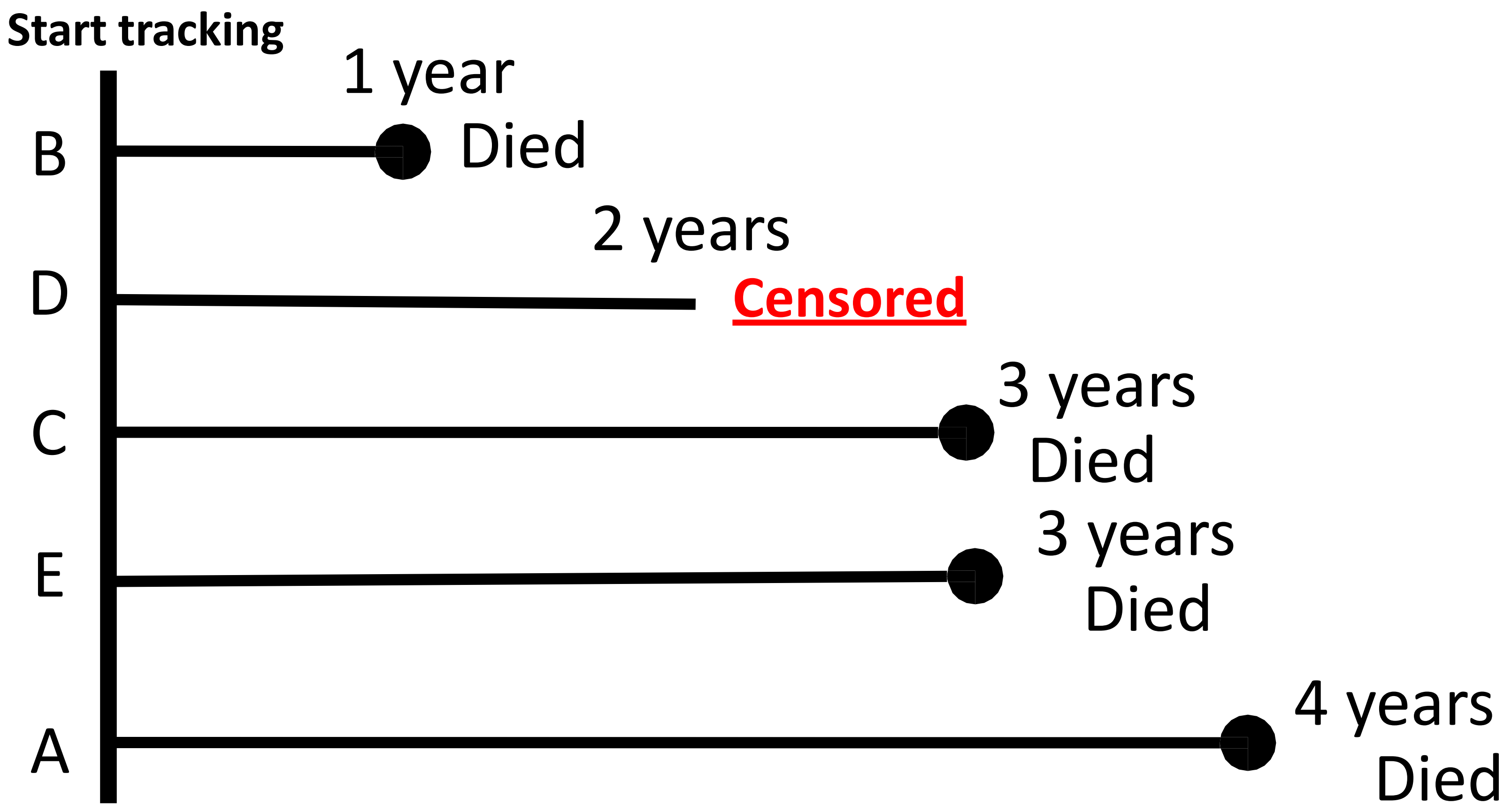
Example 2: When There is Censoring

- Estimation of the survival curve using the Kaplan–Meier method



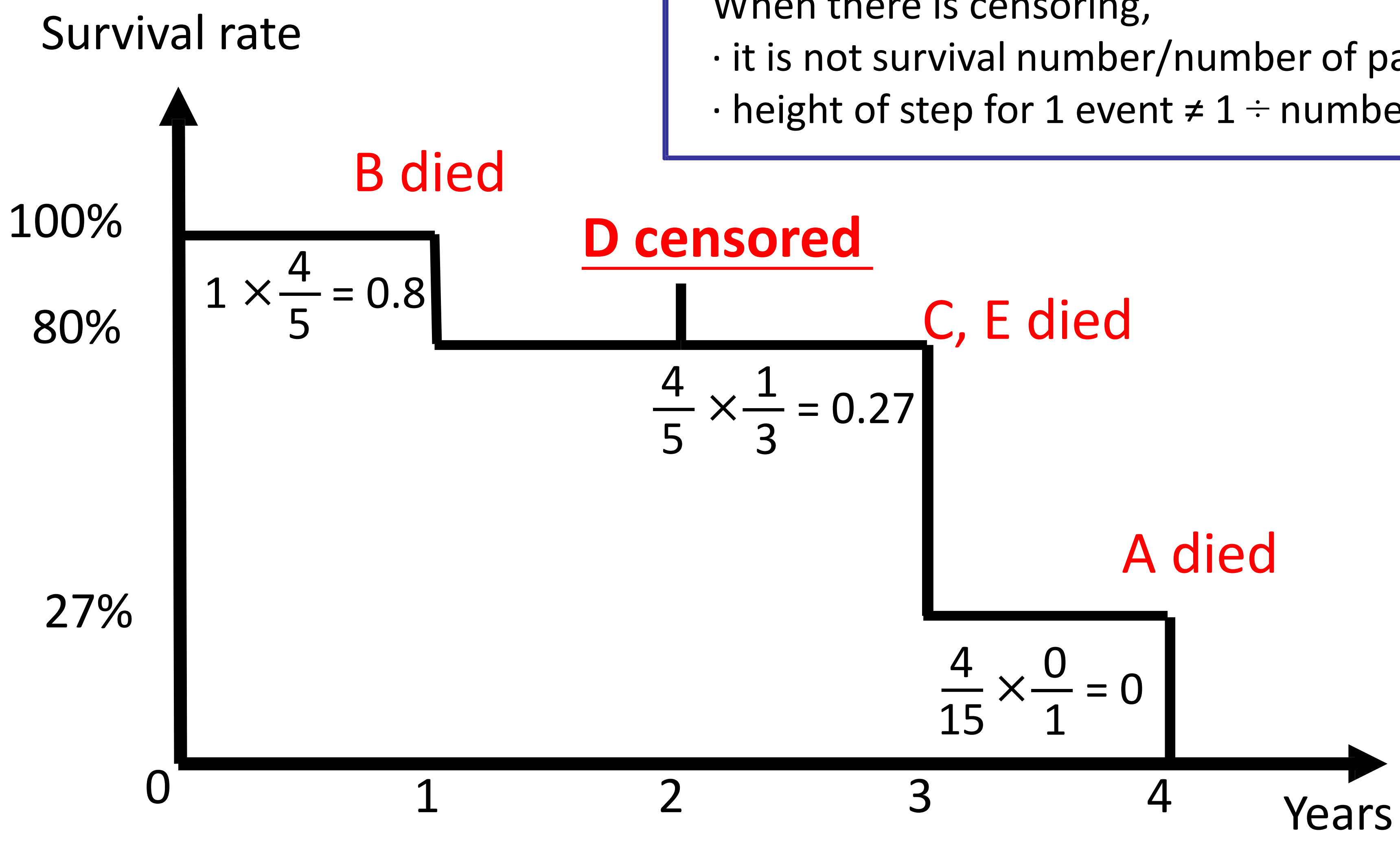
Rearrange in the Order of the Shortest to Longest Survival

- In short order, B → D → C, E → A

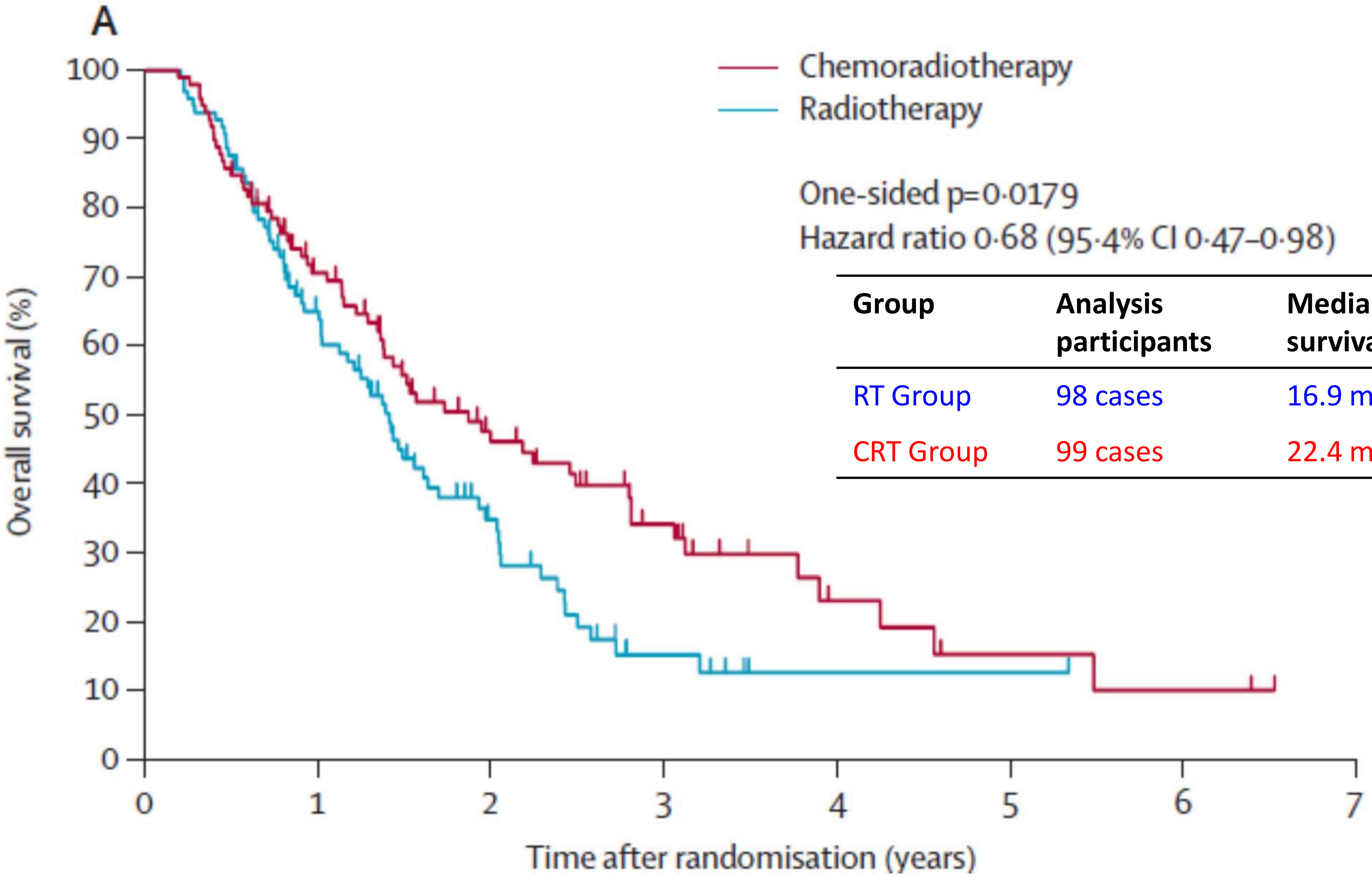


Kaplan–Meier Curve (KM Curve)

When there is censoring,
· it is not survival number/number of participants
· height of step for 1 event ≠ 1 ÷ number of participants



Determine the Values in the Table and the Survival Curves Using the Kaplan–Meier Method



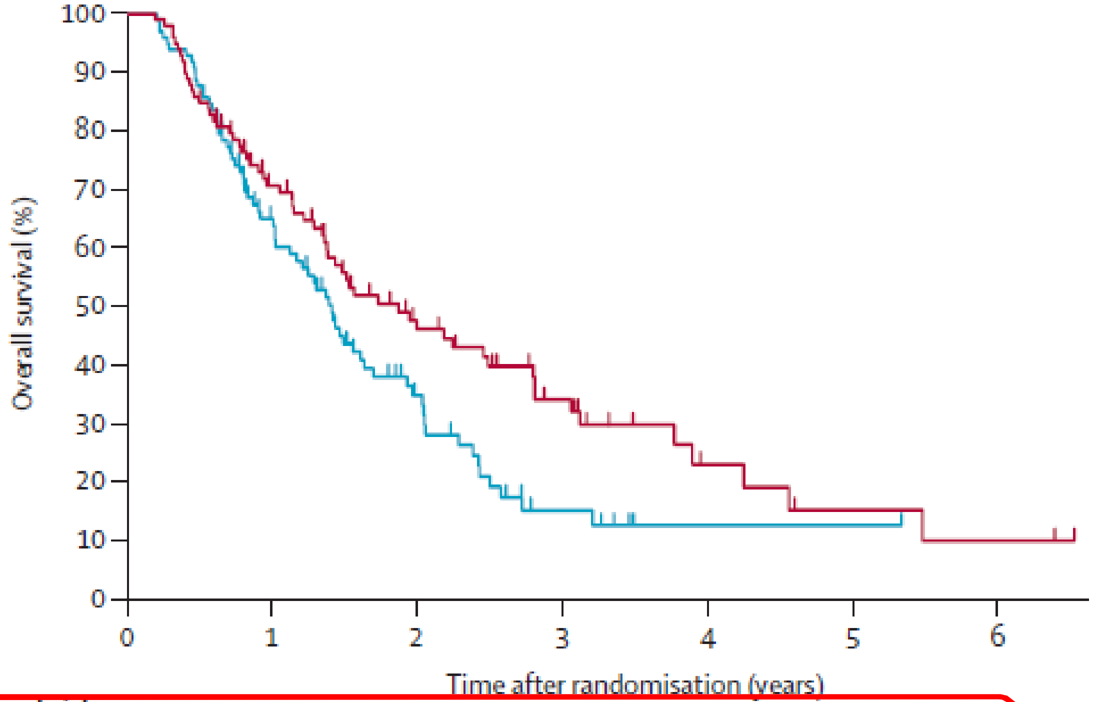
Group	Analysis participants	Median survival	2-year survival rate
RT Group	98 cases	16.9 months	35.1%
CRT Group	99 cases	22.4 months	46.3%

Number at Risk

Number [Patients] at Risk

- The number of cases that have been followed up to a certain time point (without receiving an event or censored)
- Recently, we are often asked by journals to describe "if the tracking is sufficient"

Number at risk

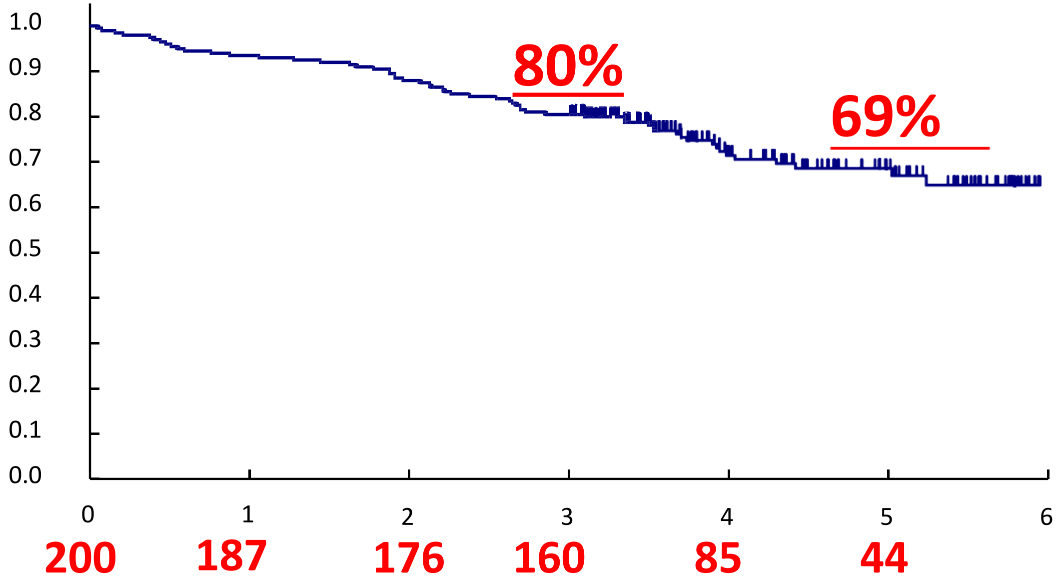


Number at risk							
Chemoradiotherapy	99	60	31	17	6	3	2
Radiotherapy	98	54	21	6	1	1	0

Atagi *et al.* (2012) Lancet Oncol. 13(7): 671-8.

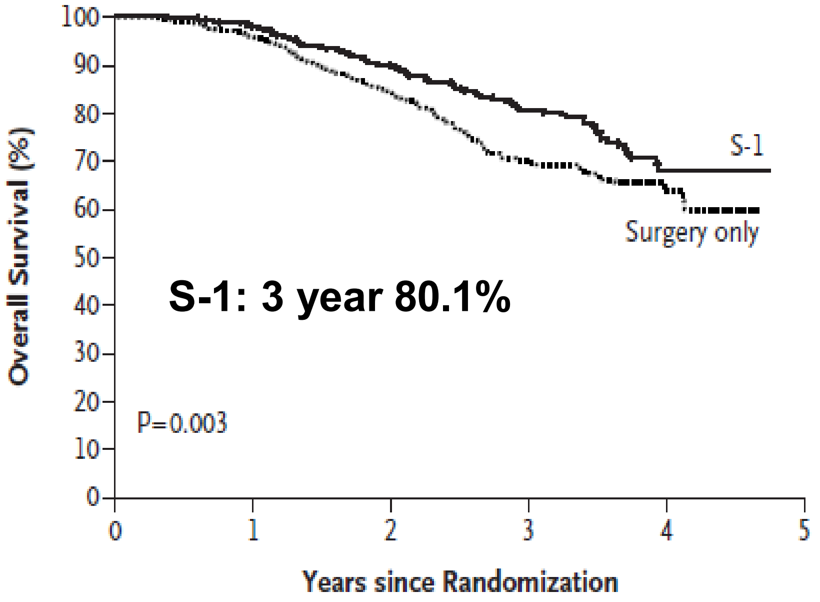
Predictable Distribution of Censoring from Number at Risk

- If there is no censoring, the survival rate of KM curve
  = survival number/number of participants
- At 3 year
  - 160/200 = **80%**
  - Match with KM curve
- At 5 year
  - 44/200 = **22%**
  - Different from KM curve
  - Many censorings between 3 and 5 years



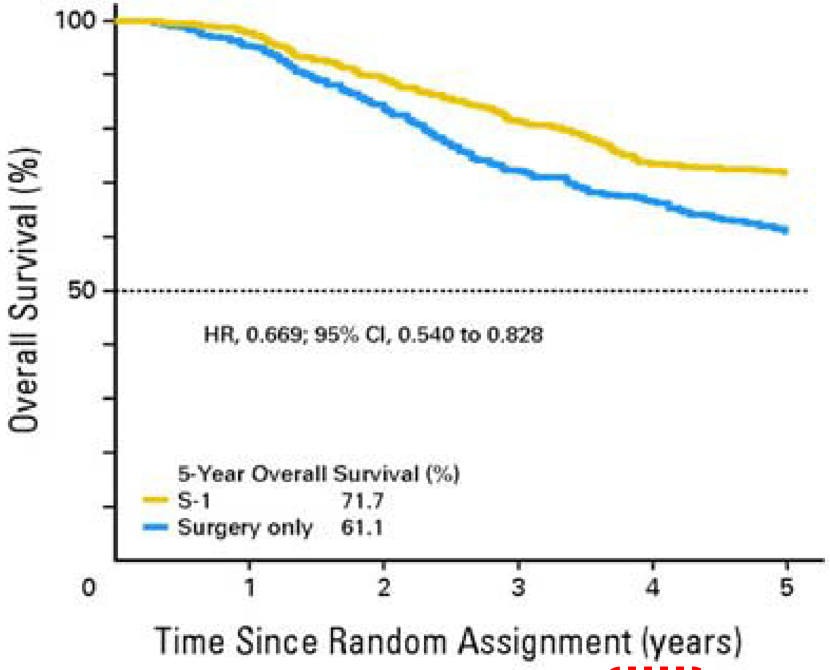
The Study Accuracy Can be Judged Without a Censoring Sign

- Compare numbers at risk to see how well they have been tracked.
  - S-1 group at 4 years at risk 46 ⇒ 363



No. at Risk					
S-1	529	515	370	196	46
Surgery only	530	504	352	163	40

Sakuramoto S et al. NEJM **2007**; 357: 1810-20



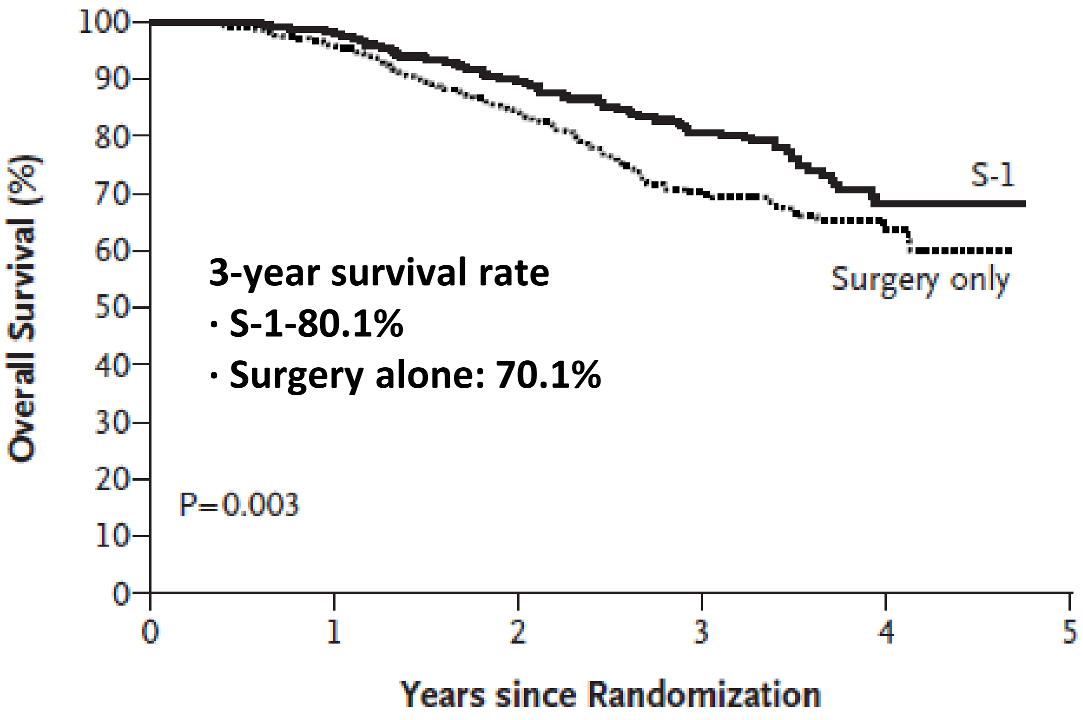
No. at risk						
S-1	529	515	465	416	363	316
Surgery only	530	504	438	365	327	268

Sasako M et al. JCO **2011**; 29: 4387-4393

Log-Rank Test

At What Time Point Is the P value of the Log-rank Test Compared?

- Protocol Setup
  - **5-year** survival rate 80% vs. 70% (HR = 0.70)
  - Enrollment 3 years, follow-up **5 years**
- Results at 1 year and 6 months after the end of registration
  - Log-rank P = 0.003
  - **3-year survival rate**
    - 80.1% vs. 70.1%



Sakuramoto S et al. NEJM 2007; 357: 1810-20

Which Understanding Is Correct?

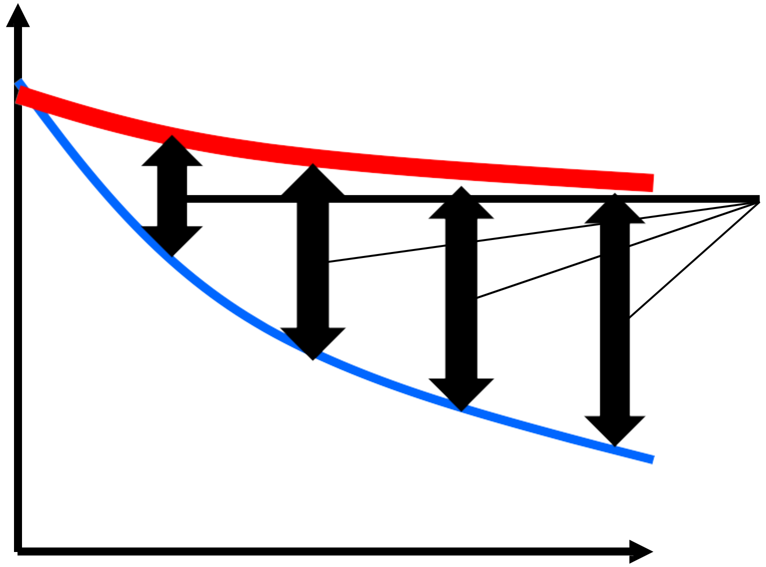
Because the protocol states that the follow-up period is 5 years, I think this is a comparison of the **5-year survival rate**.
But why does it show the 3-year survival rate?

In advanced cancer, we usually compare MSTs; thus, I think this is also a comparison of **MSTs**.
However, this has not reached MST.
How was it calculated?

I think this is a comparison of the **3-year survival rate**, as 3-year survival rate is written. The curve is the widest around 3 years, and I think that area was intentionally chosen, wasn't it?

The Key Idea of the Log-Rank Test (1)

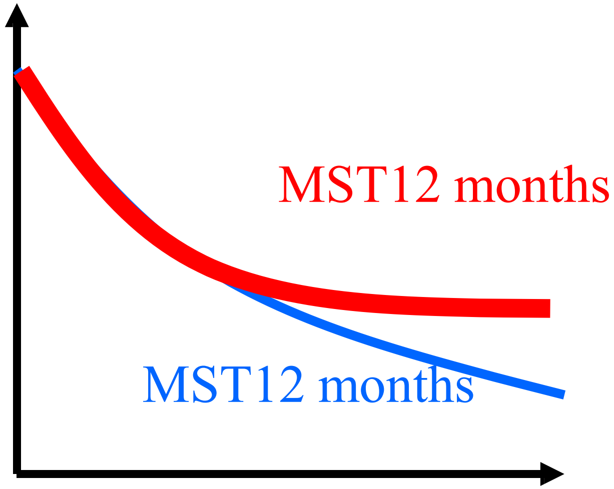
- The entire survival curves are compared
  - It does not compare the differences at a particular time point!



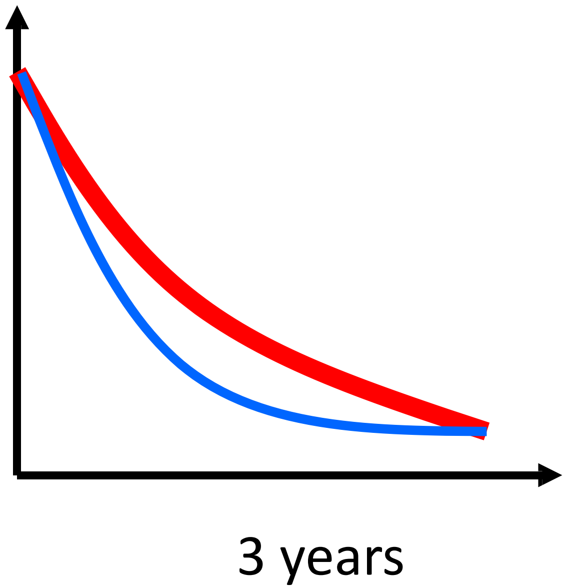
- The differences across the entire curve are assessed by summing the gaps in the survival curve at each time point when events occur.
- In other words, the further apart the two survival curves are, the larger is the gap (= lower *P* value).

Because It Is Not the Difference at One Time Point ...

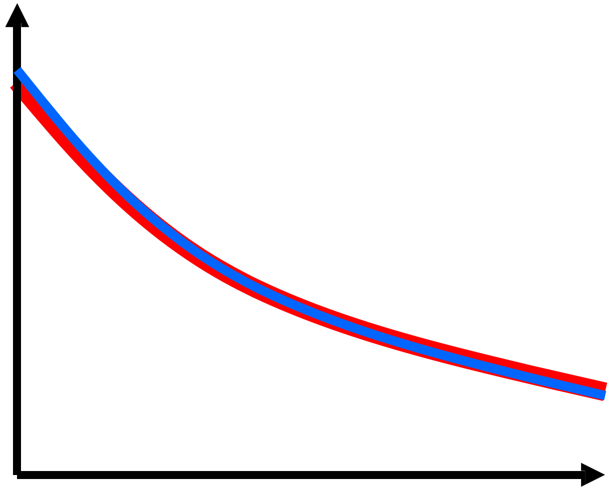
Even MST is the same,
two-sided p ≠ 1.00



Even if the curves
overlap at 3 years,
two-sided p ≠ 1.00

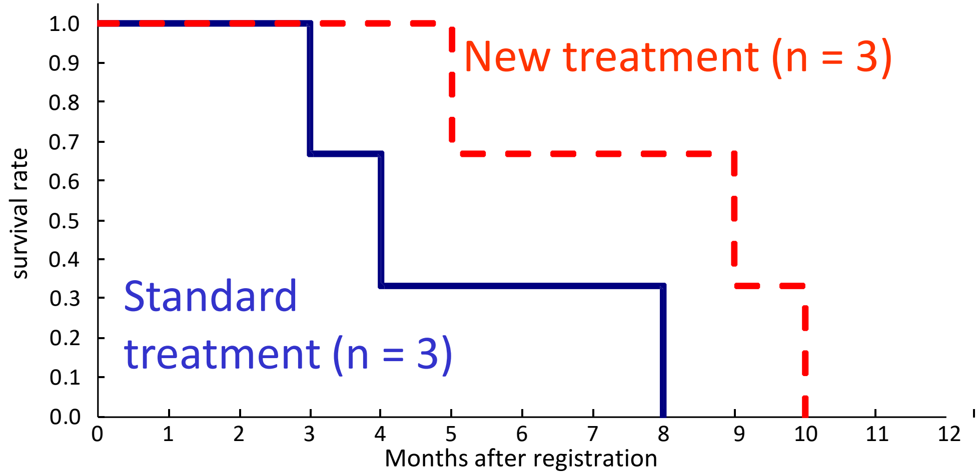


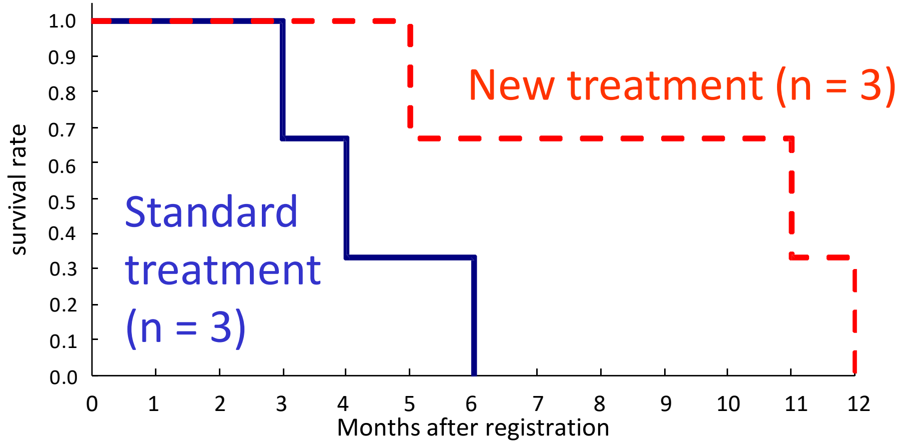
If they overlap all the way,
two-sided p = 1.00



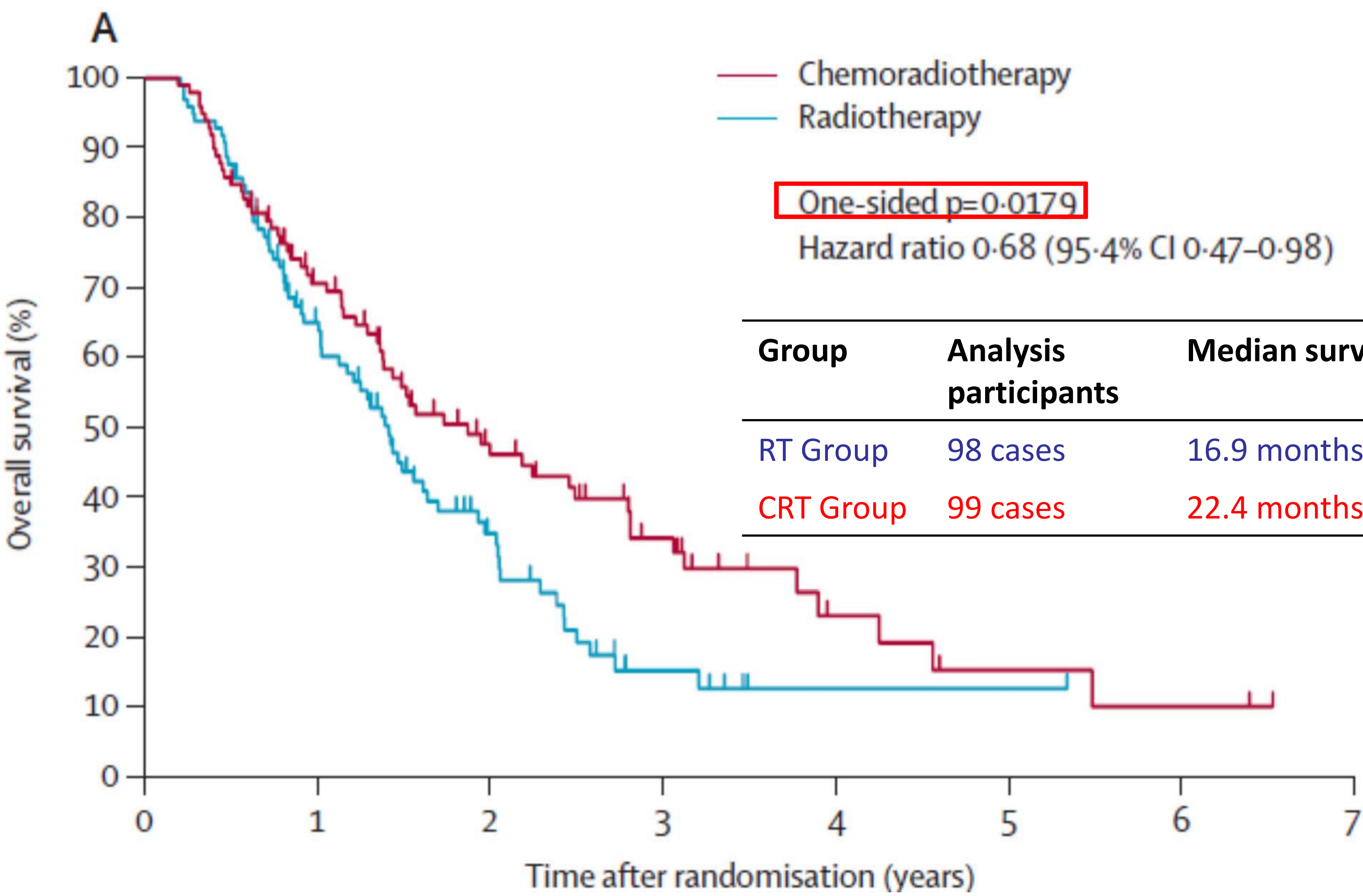
The Key Idea of the Log-Rank Test (2)

- **Only the order** in which **events** occurred is used in the calculation of the test
  - No calculation for no event
  - **Time** when the event occurred is not used.
- The following comparisons of the curves have the same log-rank test *P* value
  - The same idea applies to the Cox regression described later. Both have the same hazard ratio (HR) and 95% confidence interval (CI)
  - Two-sided log-rank: p = 0.108, HR 0.19, 95% CI [0.02–1.84]





The *P* Value is Obtained Using the Log-Rank Test



Group	Analysis participants	Median survival	2-year survival rate
RT Group	98 cases	16.9 months	35.1%
CRT Group	99 cases	22.4 months	46.3%

Cox Regression

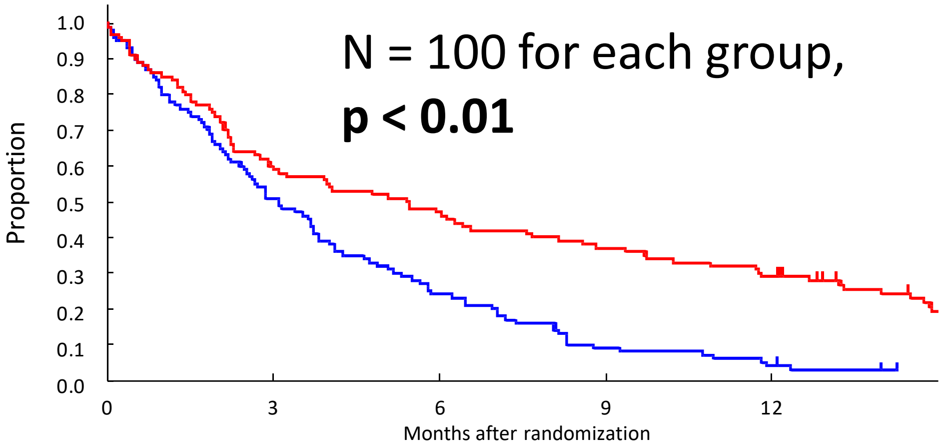
(Cox Proportional Hazards Model)

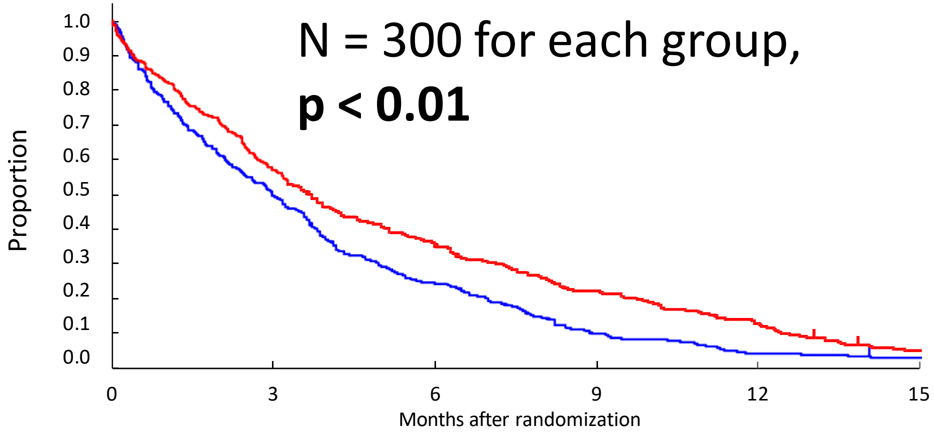
Statistical Significance ≠ Clinical Significance

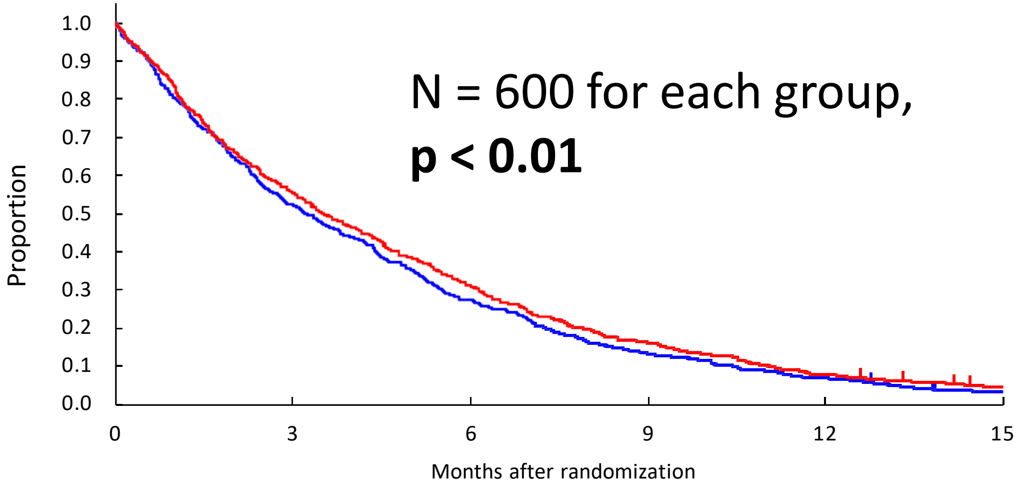
The same $p < 0.01$ could have different clinical meanings

P **value is not an indicator of the magnitude of treatment effect**

Note: Hypothetical example





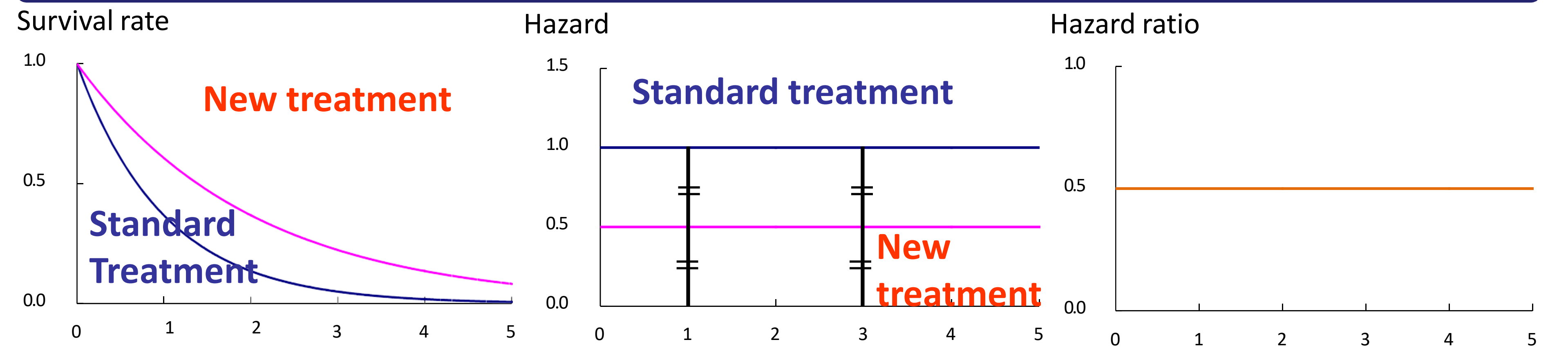


Indicator of the Magnitude of Treatment Effect

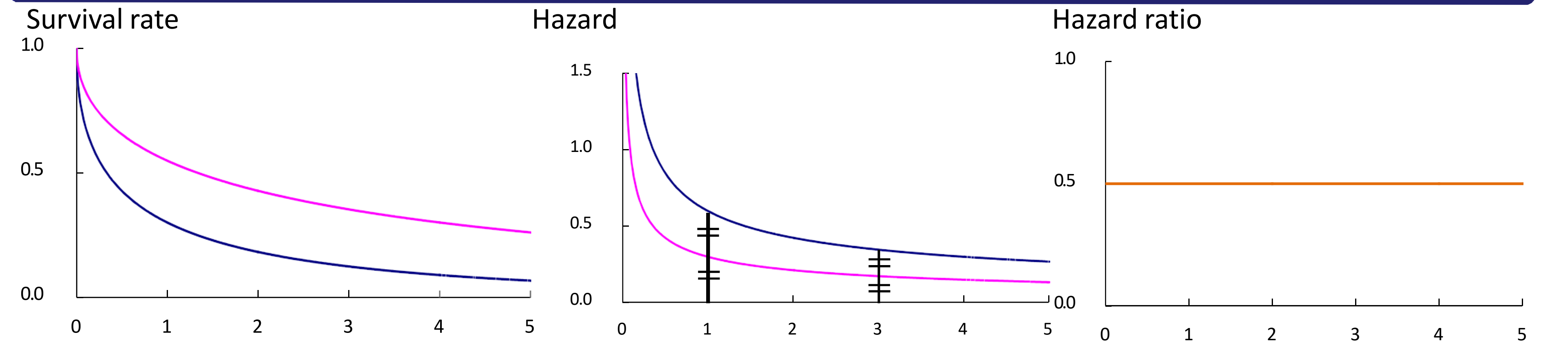
- Indicator focused on one point on the curve
  - Annual survival rate
    - 2-year survival rate CRT: 46.3% vs. RT: 35.1%
  - Median Survival Time (MST)
    - CRT: 22.4 months vs. RT: 16.9 months
- Indicator that combines the entire curve into one effect
  - **Hazard Ratio (HR)**
    - **Hazard (instantaneous mortality)** ratio between groups

Relationship between “Survival Curve,” “Hazard,” and “Hazard Ratio”

Exponential distribution: Hazard is constant through time if both groups have exponential distribution, HR is constant

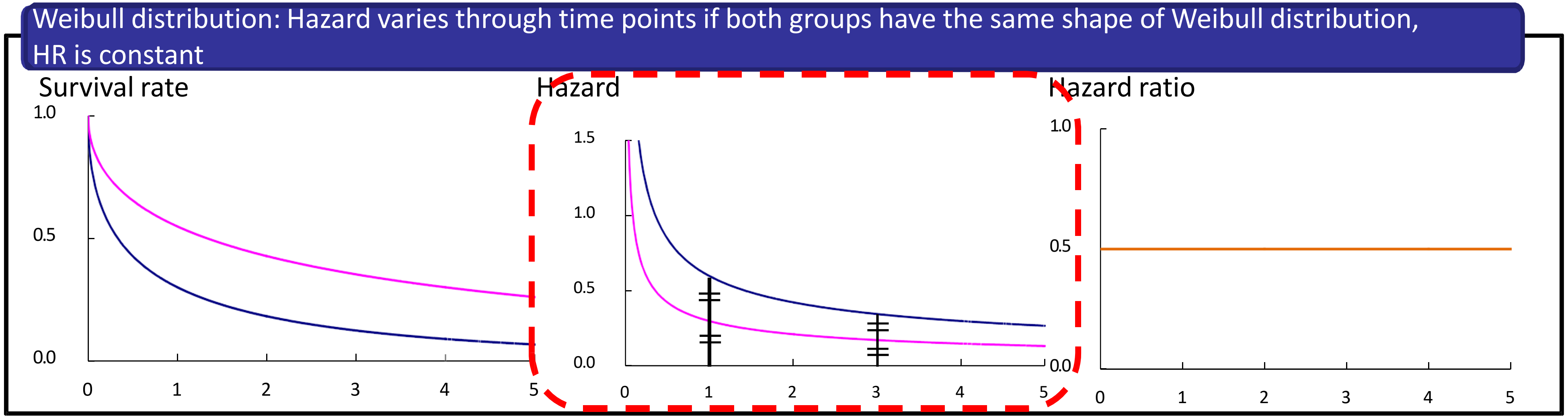


Weibull distribution: Hazard varies through time points if both groups have the same shape of Weibull distribution, HR is constant



When the HR is constant regardless of the time point, “**proportional hazard** is established.”

Cox's Proportional Hazards Model (Cox Regression)

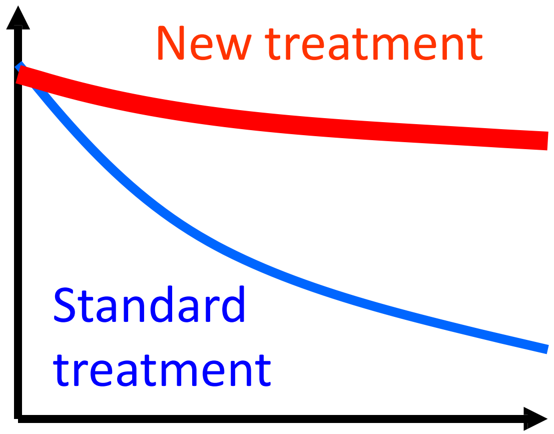


- The model represented by an equation as shown below
  - models the relationship between time and the hazard at that time point

Hazard for the new treatment group at time point t
= Hazard for standard treatment group at time point t × exp(β)

- For multivariate analysis, multiple variables are added in exp

What You Can Do with Cox Regression



- The magnitude of the treatment effect can be estimated
  - Summary Indicator: **Hazard ratio (HR)**
- When determining the HR of the new treatment group relative to the standard treatment group,
  - hazard of standard treatment cases
    = (hazard for standard treatment group at time point t) × 1
  - hazard of new treatment cases
    = (hazard for standard treatment group at time point t) × exp(β)

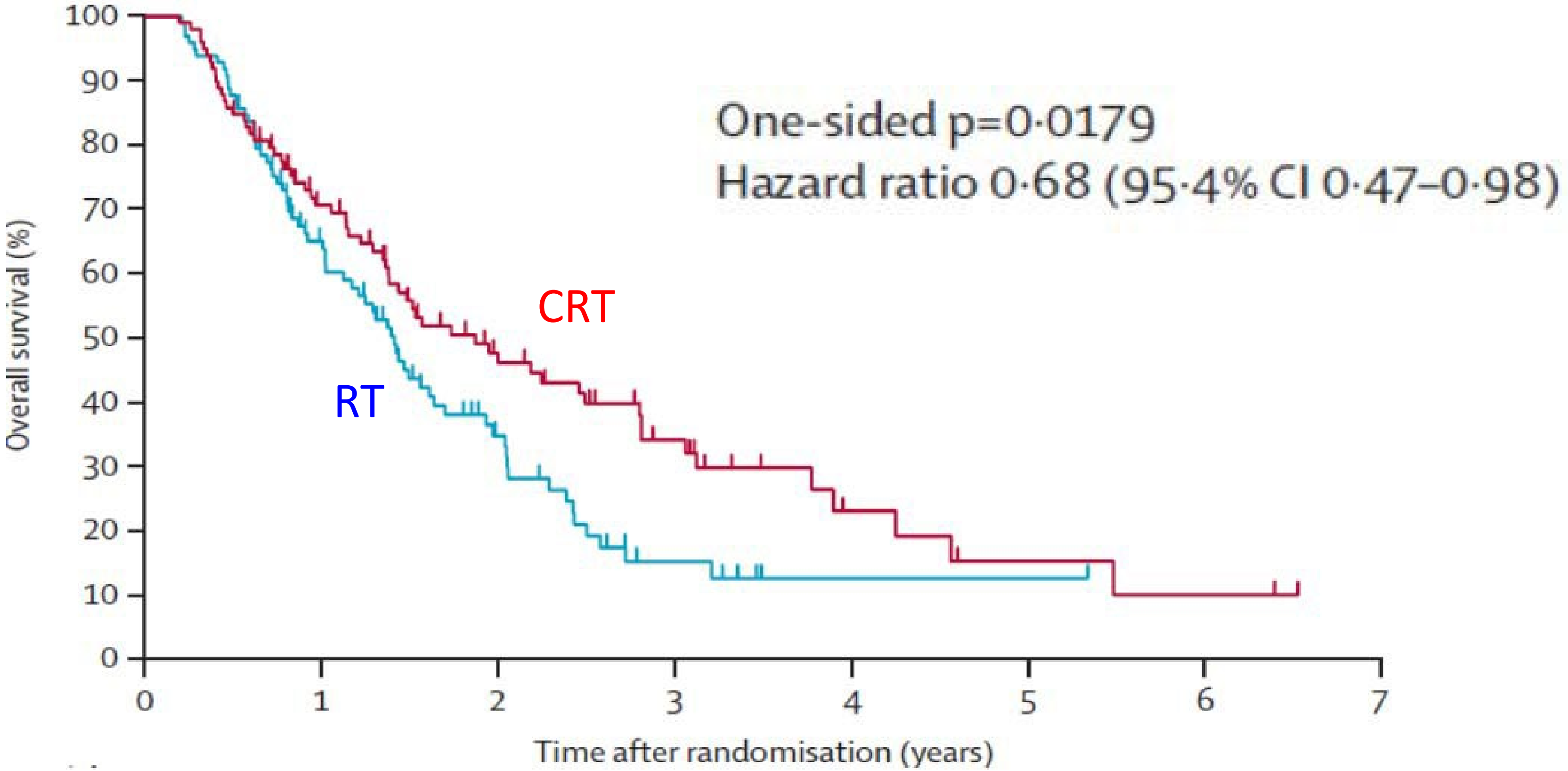
$$\text{HR} = \frac{\text{Hazard for new treatment group}}{\text{Hazard for standard treatment group}} = \frac{\cancel{\text{Hazard for standard treatment group}}}{\cancel{\text{Hazard for standard treatment group}}} \cdots \frac{\exp(\beta)}{1} = \exp\boxed{(\beta)}$$

Estimation from data

Interpretation of Hazard Ratio (HR)

- HR of CRT for RT = 0.68
  - CRT reduces risk of death by 32% over RT.
  - CRT is 0.68 times more riskier than RT.

Atagi *et al.* (2012) Lancet Oncol. 13(7): 671-8.

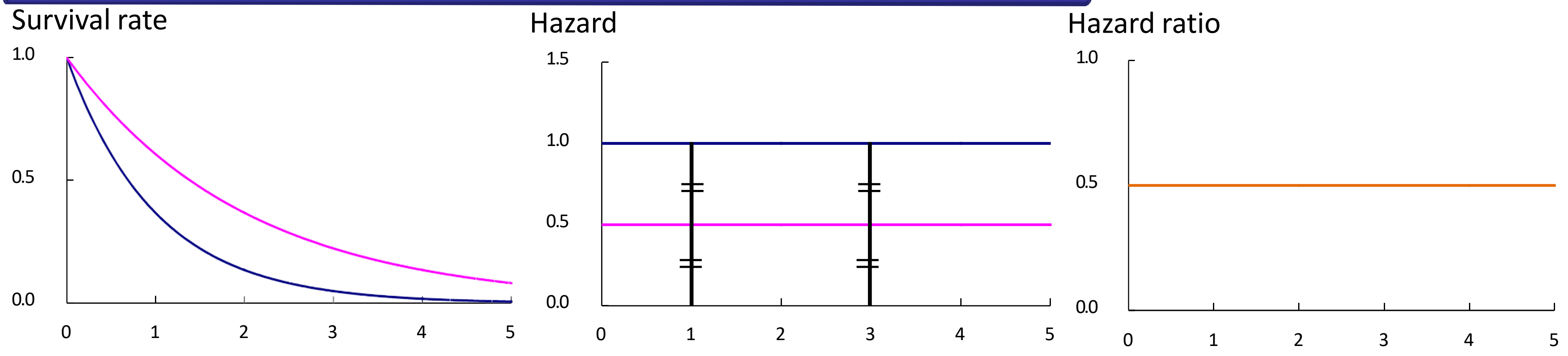


Assumptions of Cox Regression

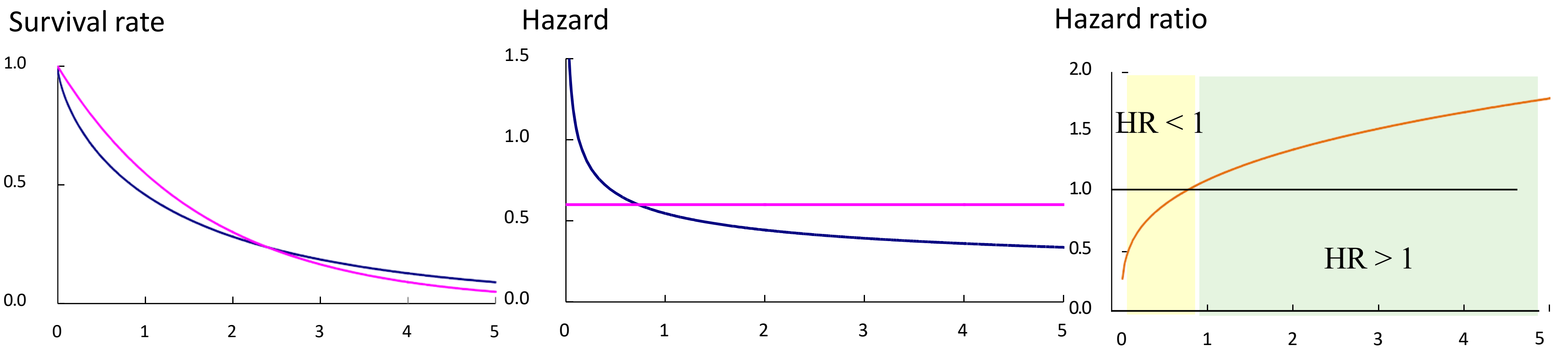
- There is one HR that can be obtained by Cox regression
  - Averages the effects of the entire curve and summarizes them into one
- The “proportional hazard property” holds true
  - “Hazard proportionality is almost constant at all time points”
- If the proportional hazard does not hold true,
  - it is not appropriate to use a single HR to summarize the effect
  - interpretation becomes more difficult

Validity of HR as determined by Cox regression

Proportional hazard holds true. HR = 0.5 obtained by Cox regression.



Proportional Hazard does not hold true. HR = 1.2 obtained by Cox regression ⇐ Interpretable?



HR is not appropriate when **proportional hazard** does not hold true.

IPASS study

Mok, Tony S., et al. *NEJM* 361.10 (2009): 947-957.

Stage IV/recurrent Non-small cell lung cancer 1st line

Random assignment

608 Examples

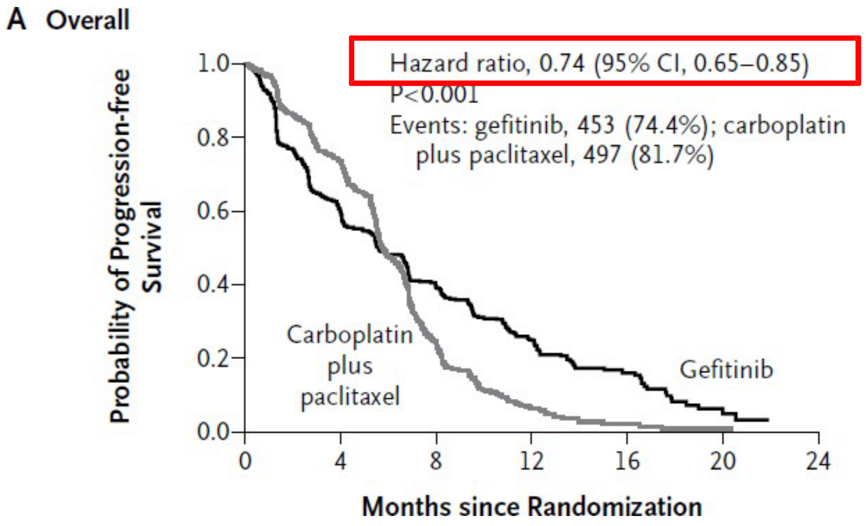
Carboplatin+ Paclitaxel

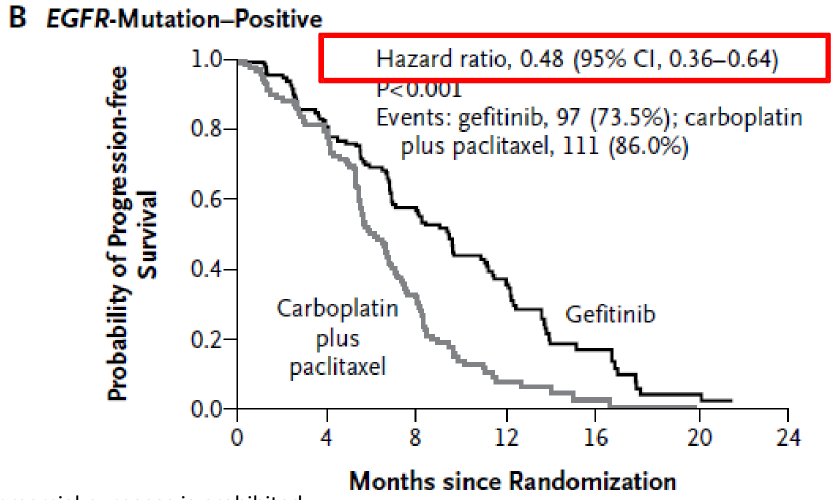
(Standard treatment)

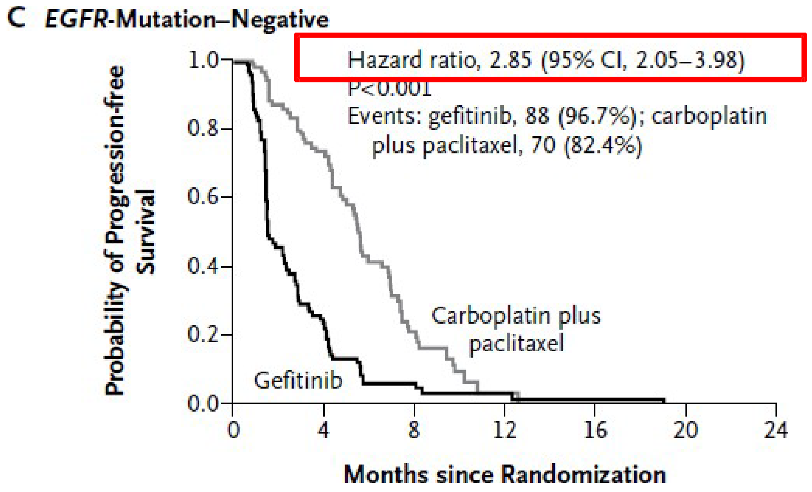
609 Examples

Gefitinib

(New treatment)







Dutch D1D2 trial

Songun, Ilfet, et al. *Lancet Oncol.* 11.5 (2010): 439-449.

Resectable stomach cancer

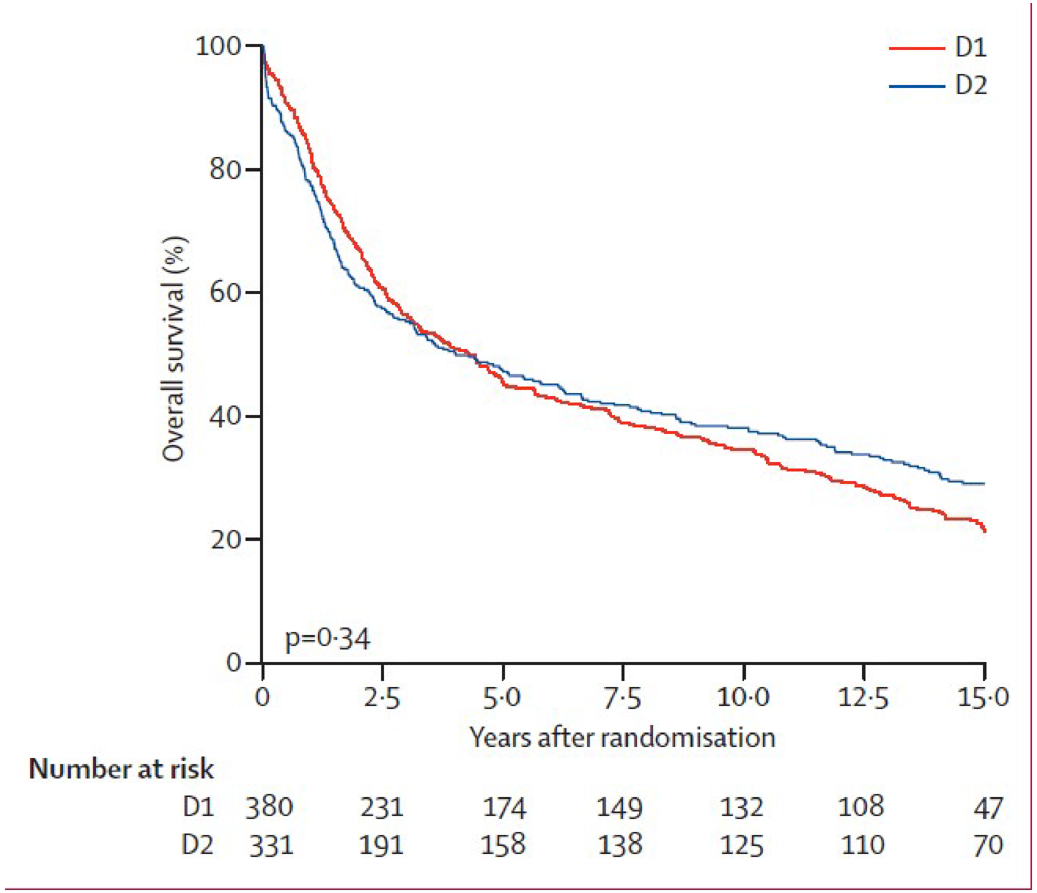
↓

Random assignment

↓

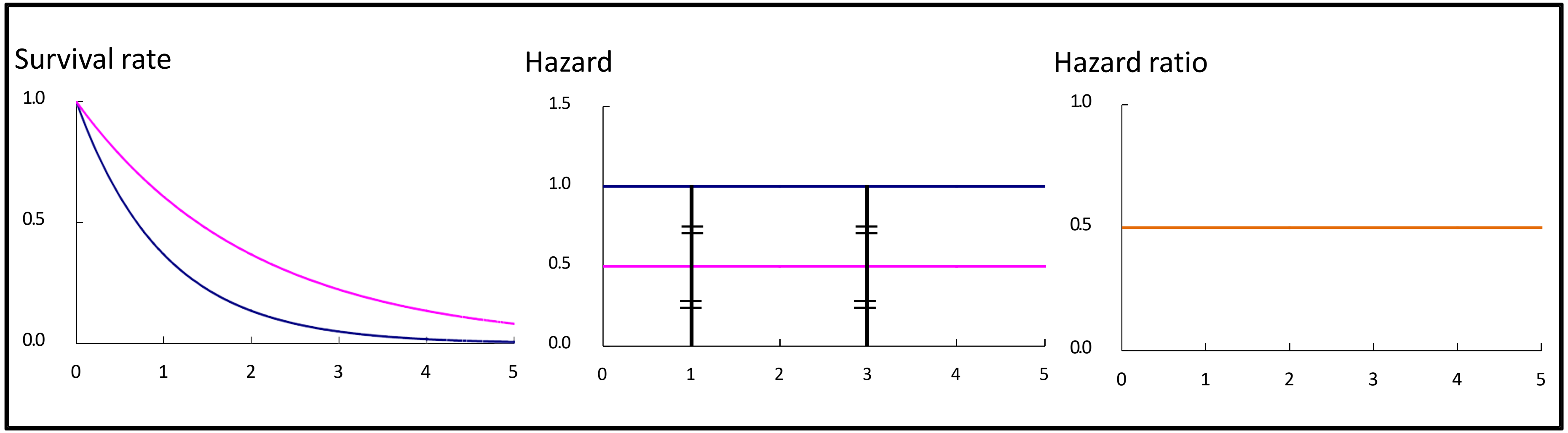
D1 dissection (Standard treatment) | **D2 dissection** (New treatment)

	D1	D2
Postoperative death	15 (4%)	32 (10%)
Complications (in an illness)	94 (25%)	142 (43%)
Repeat surgery	30 (8%)	59 (18%)



Difference between Rate, Ratio, and Proportion

Definition of Usage Terms



- Survival rate <u>sometimes referred to as "survival proportion"</u>
- Can mortality <u>**be**</u> calculated as "1 - survival rate"?
- To which concept does "hazard" apply?

Proportion, Rate, and Ratio: From Kojien (Japanese Dictionary)

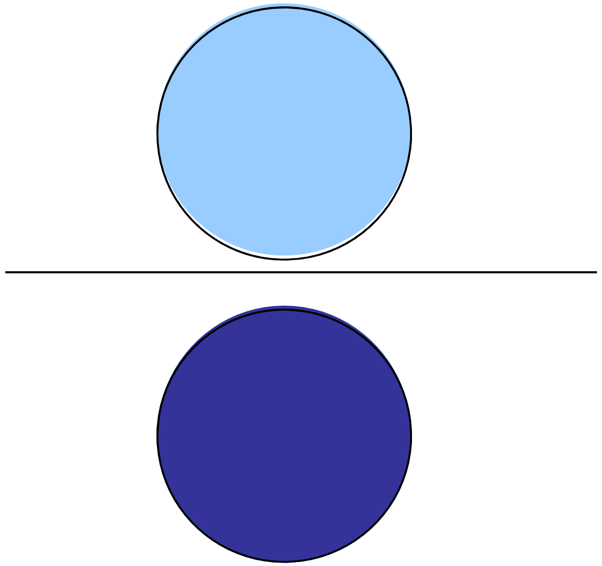
- Rate: proportion, percentage
- Proportion: ratio between quantities, percentage, ratio.
- Ratio: two quantities A and B of the same kind exist, and when B is not zero, the relation of how many times A is equal to B is called the proportionality of A to B. This is written as A:B.

Everything becomes “comparison”

Ratio

- It has the broadest meaning among the three. It is a quantity divided by another quantity (in general, it has a dimension).

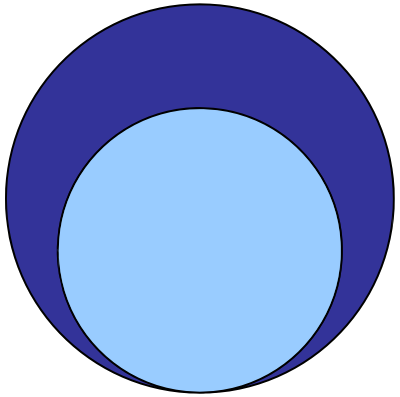
- The denominator and numerator are separate and do not contain each other's counterparts.
- Example
  - Sex ratio = number of men/number of women
  - BMI = weight (kg)/height 2 (m)



Proportion

- Form of ratio in which numerator is contained in denominator (no dimension, between 0 and 1)

- Example
  - Proportion of men = number of men/size of a population
  - Batting average = hits/at bat
  - Response rate = (CR + PR)/all eligible counts



Rate

- Relates to the rate of change of a phenomenon under certain conditions (has dimension, may exceed 1)

- Example
  - Divorce rate = number of divorces **per year** per 1,000 people
  - Mortality (Hazard by person-year method)
    - Number of deaths/Total observation time

      = 2 deaths per 3 person-years

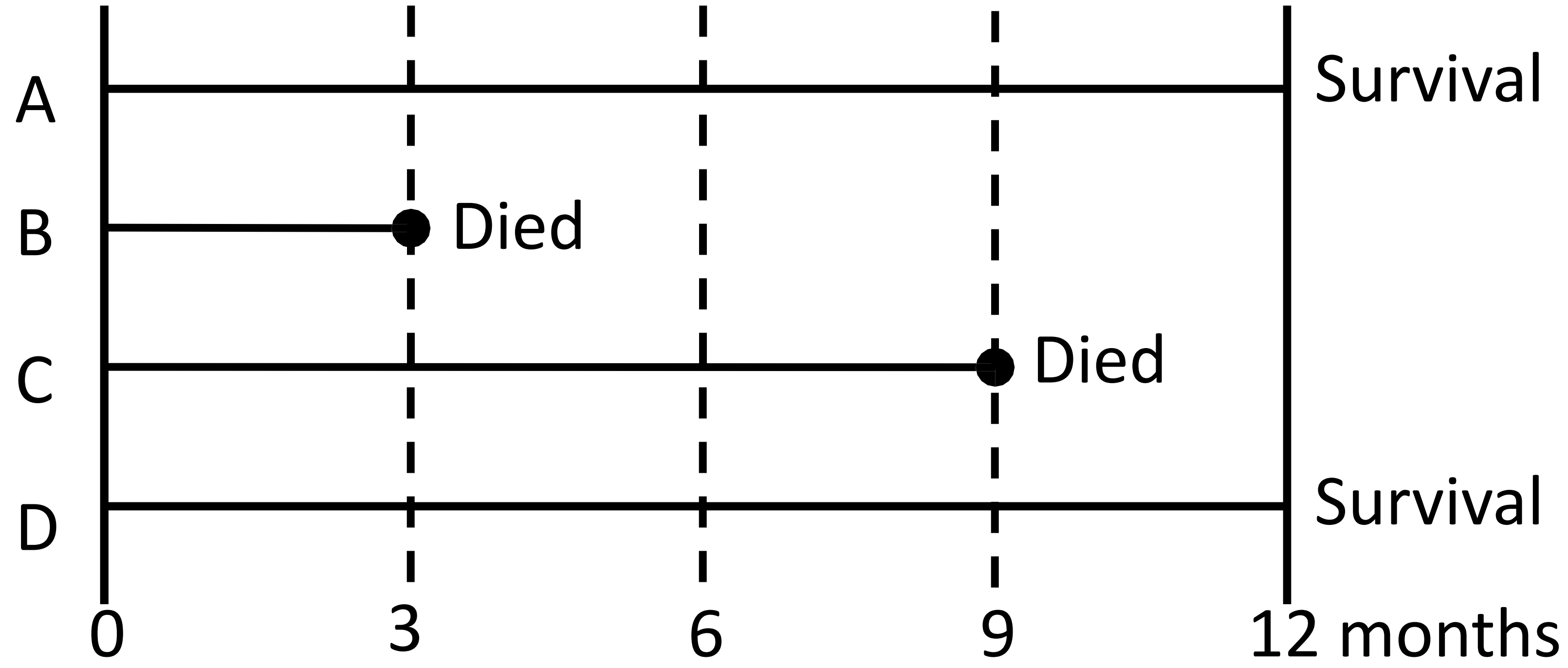
      = 66.7 deaths per 100 person-years

  ⇔ Interpretation: If 100 people are observed for 1 year, 66.7 will die.

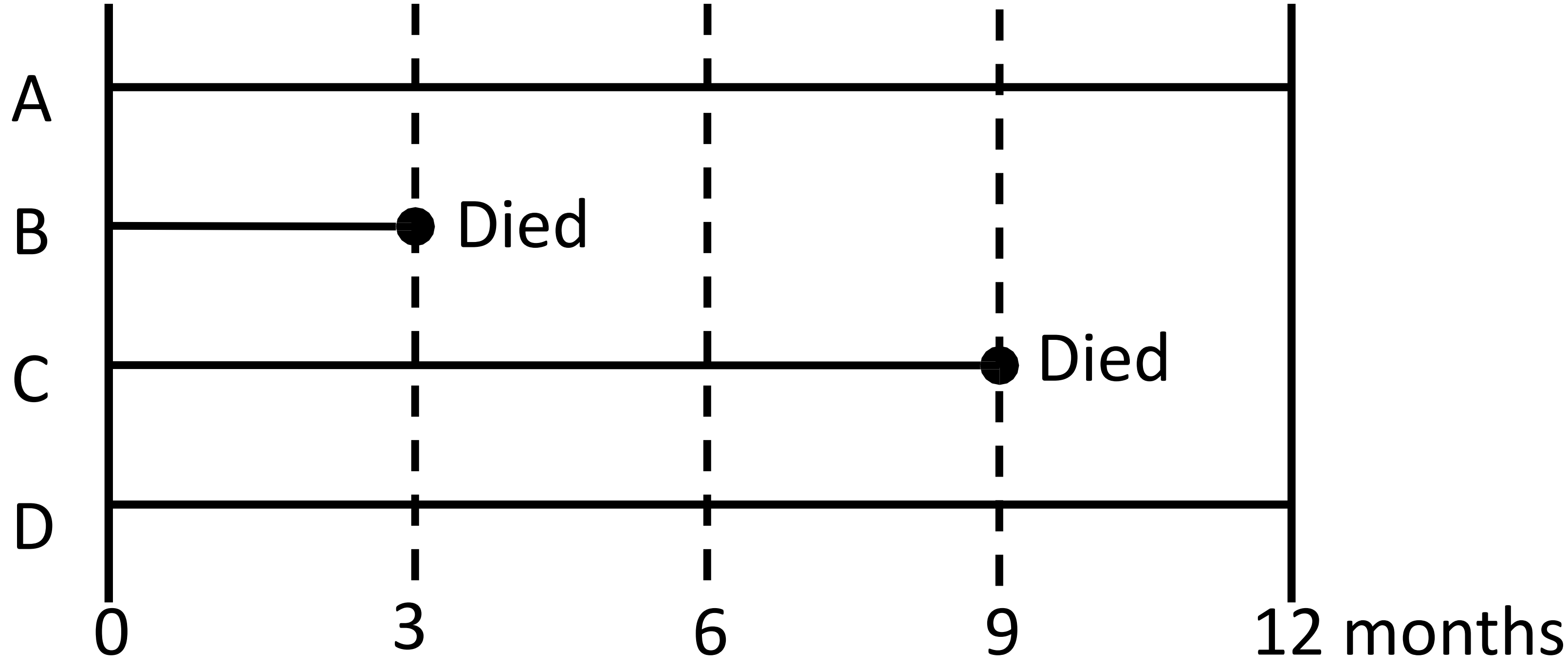
Question 1: Which One Is It among Rate, Ratio, and Proportion?

1. Batting average: **proportion**
2. Mortality: **rate (Mortality ≠ 1 − survival proportion!)**
3. Response **rate**: **proportion**
4. Survival **rate**: **proportion**
5. Hazard: **rate**

Question 2. What Is the Proportion of Deaths and Mortality for These Four Individuals?



Answer to Question 2



Proportion of deaths: 2 people/4 people = 0.5

Mortality: 2 people/3 years = 0.67 people/year

2 people/36 months = 0.056 people/month

Outcome Types and Statistical Methods

	Continuous Variable	Binary (0/1)	Survival Time
Outcome examples	BMI, blood pressure, laboratory values	Response rate	Overall survival, progression-free survival
Data summary	Histogram	Contingency table	**Kaplan-Meier method**
Group comparison (test)	*t*-test, Wilcoxon test	Chi-square test, Fisher’s exact test	**Log-rank test**
Model fitting	Multiple regression analysis	Logistic regression	**Cox regression**

Summary

- To estimate the survival curve, use the **Kaplan–Meier method**.
- The survival rate determined using the Kaplan–Meier curve is the same as the number of survivors divided by the number of participants, if there is no censoring.
- **Use log-rank test** to perform an intergroup comparison of the survival curve.
- The log-rank test compares the entire curve, not just one time point
- **Use Cox regression** to obtain the treatment effect
- The treatment effect can be summarized by hazard ratio based on the Cox regression.
- When the proportional hazard does not hold true, hazard ratio determined by Cox regression is not appropriate.
- In research, be aware of the difference between rates, ratios, and proportions.