Biostatistics

Chapter 10 Survival Analysis

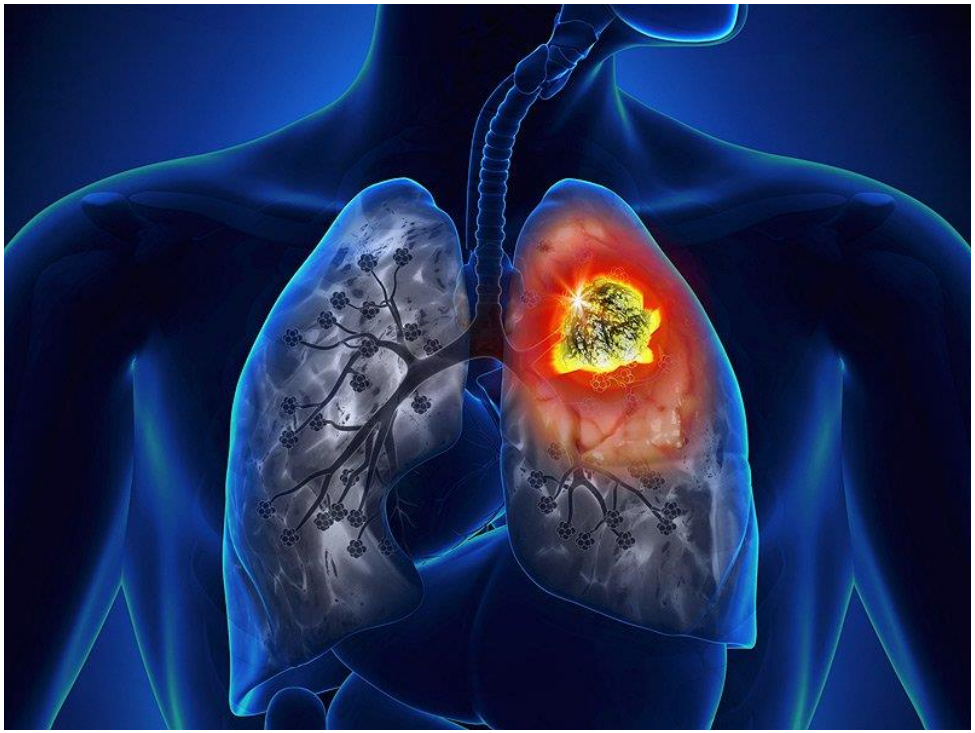
Jing Li

jing.li@sjtu.edu.cn

http://cbb.sjtu.edu.cn/~jingli/courses/2017fall/bi372/

Dept of Bioinformatics & Biostatistics, SJTU

Example: lung cancer



The prognosis for NSCLC (non-small cell lung cancer) patients remains poor, with the 5-year overall survival (OS) rate of 15% of all stages.

Cancer statistics, 2013.

CA Cancer J Clin. 2013 Jan;63(1):11-30.

Example: lung cancer -immunotherapy

FDA U.S. FOOD & DRUG ADMINISTRATION

A to Z Index | Follow FDA | En Español

Search FDA

Home | Food | Drugs | Medical Devices | Radiation-Emitting Products | Vaccines, Blood & Biologics | Animal & Veterinary | Cosmetics | Tobacco Products

News & Events

Home > News & Events > Newsroom > Press Announcements

FDA News Release

FDA expands approved use of Opdivo in advanced lung cancer

Opdivo demonstrates survival benefit in squamous and non-squamous non-small cell lung cancer

SHARE | TWEET | LINKEDIN | PIN IT | EMAIL | PRINT

For Immediate Release October 9, 2015

Release

The U.S. Food and Drug Administration today approved Opdivo (nivolumab) to treat patients with advanced (metastatic) non-small cell lung cancer whose disease progressed during or after platinum-based chemotherapy.

Lung cancer is the leading cause of cancer death in the United States, with an estimated 221,200 new diagnoses and 158,040 deaths in 2015. The most common type of lung cancer, non-small cell lung cancer (NSCLC), is further divided into two main types named for the kinds of cells found in the cancer – squamous cell and non-squamous cell (which includes adenocarcinoma).Opdivo works by targeting the cellular pathway known as PD-1/PD-L1 (proteins found on the body's immune cells and some cancer cells). By blocking this pathway, Opdivo may help the body's immune system fight the cancer cells. Earlier this year, the FDA approved Opdivo to treat patients with advanced *squamous* NSCLC whose disease progressed during or after platinum-based chemotherapy. Today's approval expands the use of Opdivo to also treat patients with *non-squamous* NSCLC.

"There is still a lot to learn about the PD-1/PD-L1 pathway and its effects in lung cancer, as well as other tumor types," said Richard Pazdur, M.D., director of the Office of Hematology and Oncology Products in the FDA's Center for Drug Evaluation and Research. "While Opdivo showed an overall survival benefit in certain non-small cell lung cancer patients, it appears that higher expression of PD-L1 in a patient's tumor predicts those most likely to benefit."

Inquiries

Media

Sarah Peddicord
301-796-2805

Consumers

888-INFO-FDA

Related Information

- FDA: Office of Hematology and Oncology Products
- FDA Approved Drugs: Questions and Answers
- NCI: Lung Cancer

Follow FDA

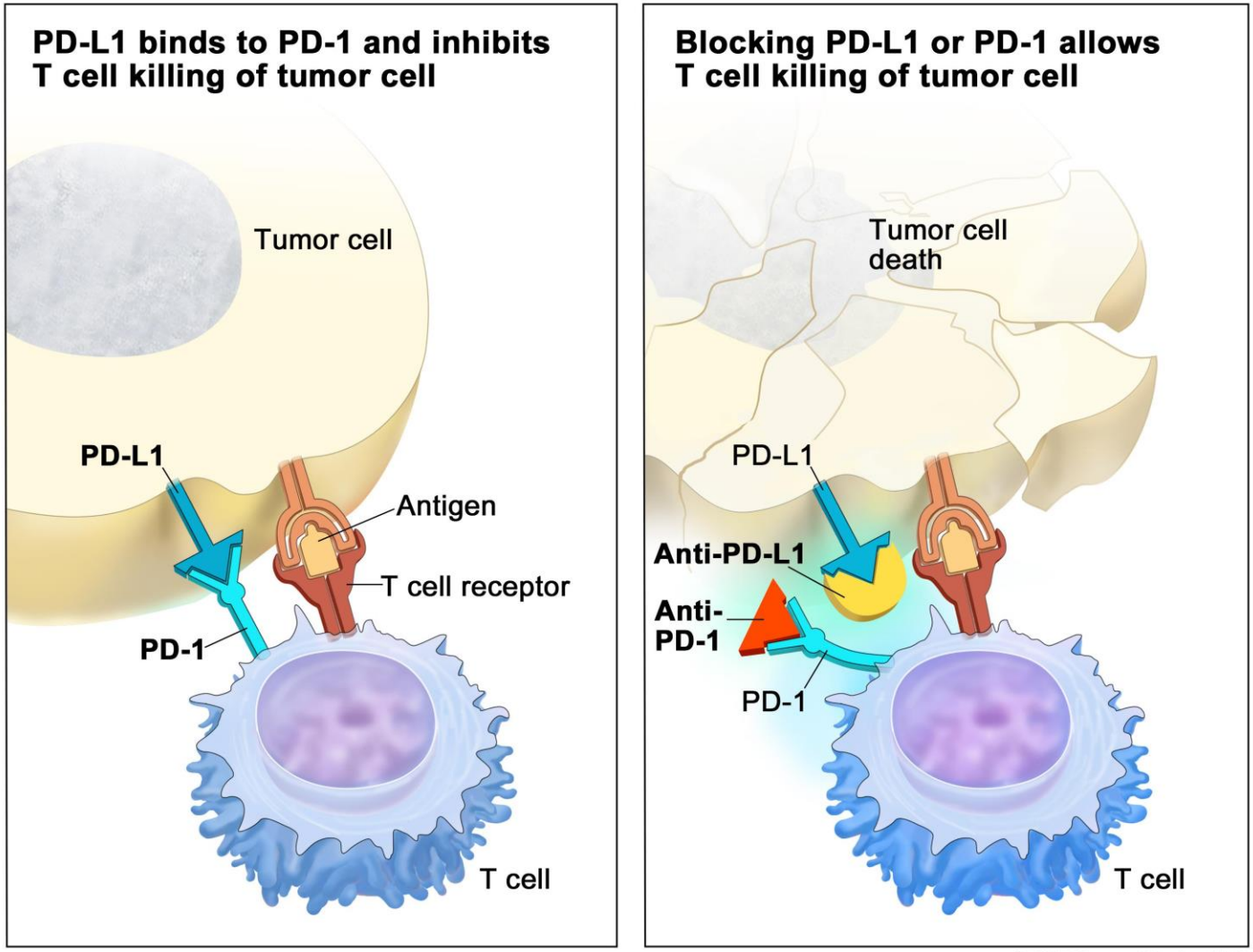
Follow @US_FDA
Follow FDA
Follow @FDAmedia

2015.10

PD-1抑制剂(Opdivo和Keytruda)默克公司

MERCK

Anti-PD-1/PD-L1 immunotherapy

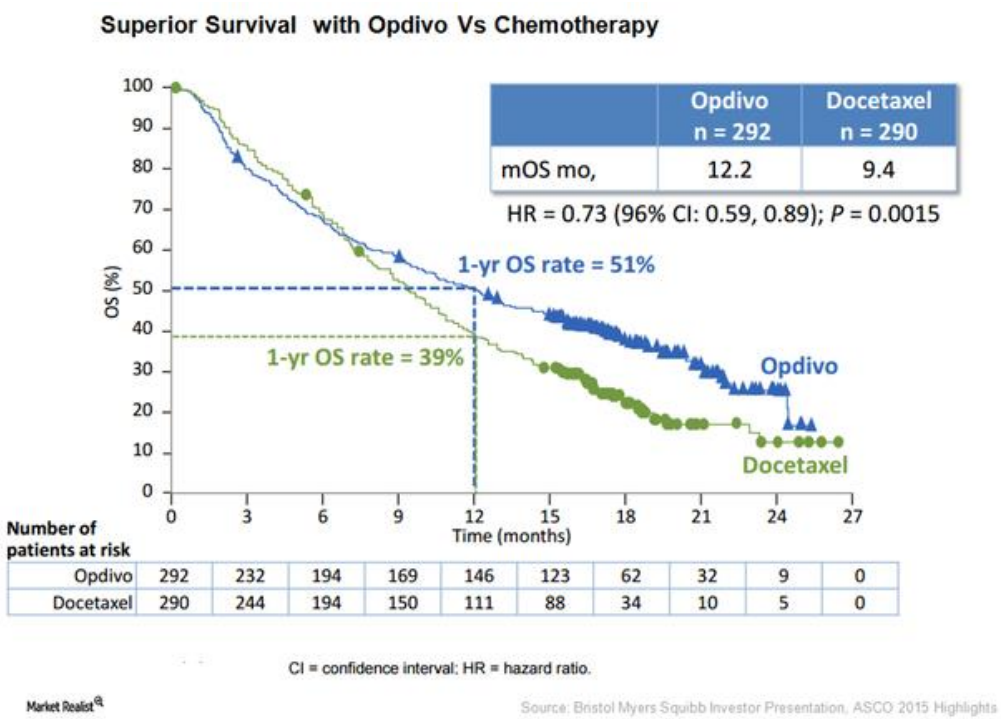


Anti-PD-1/PD-L1 immunotherapy

香港：Keytruda 100mg 32000 HKD/支；一个疗程为4支，所以目前一个疗程需要128000HKD折合人民币11万左右。

PD-1抑制剂(Opdivo和Keytruda)默克公司

Anti-PD-1/PD-L1 immunotherapy



The safety and effectiveness of Opdivo for this use was demonstrated in an international, open-label, randomized study of 582 participants with advanced NSCLC whose disease progressed during or after treatment with platinum-based chemotherapy and appropriate biologic therapy. Participants were treated with Opdivo or docetaxel (紫杉醇). The primary endpoint was overall survival, and the secondary endpoint was objective response rate (the percentage of patients who experienced complete or partial shrinkage of their tumors). Those treated with Opdivo lived an average of 12.2 months compared to 9.4 months in those treated with docetaxel.

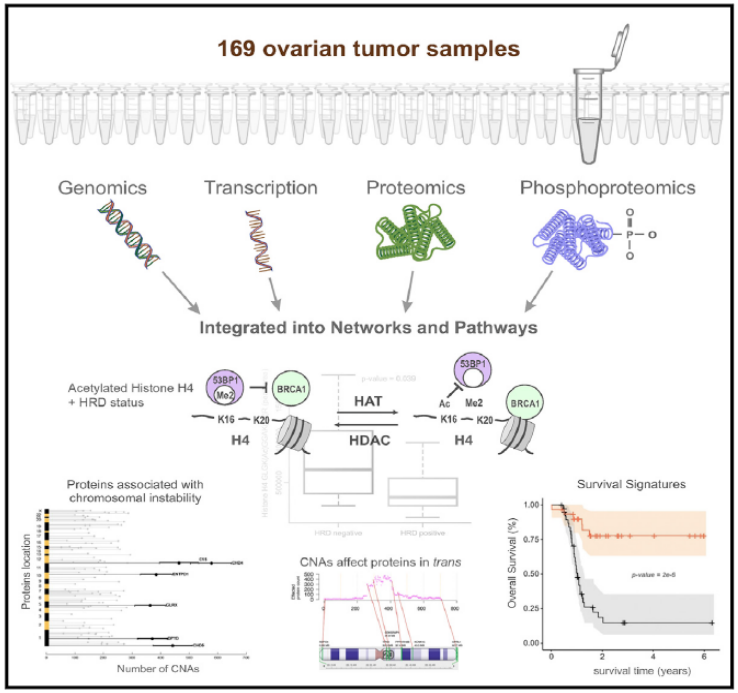
Another Example

Resource

Cell

Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer

Graphical Abstract



Authors

Hui Zhang, Tao Liu, Zhen Zhang, ...,
Daniel W. Chan, Karin D. Rodland,
the CPTAC Investigators

Correspondence

dchan@jhmi.edu (D.W.C.),
karin.rodland@pnnl.gov (K.D.R.)

In Brief

Layering proteomic and genomic data from ovarian tumors provides insights into how signaling pathways correspond to specific genome rearrangements and points to the benefit of using protein signatures for assessing prognosis and treatment stratification.

Zhang et al., 2016, Cell 166, 755–765

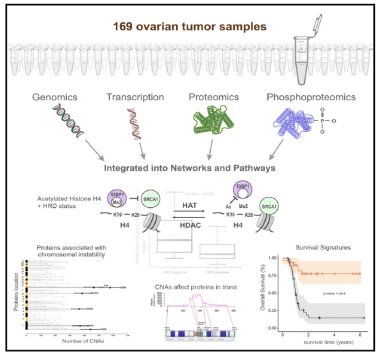
Another Example

Resource

Cell

Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer

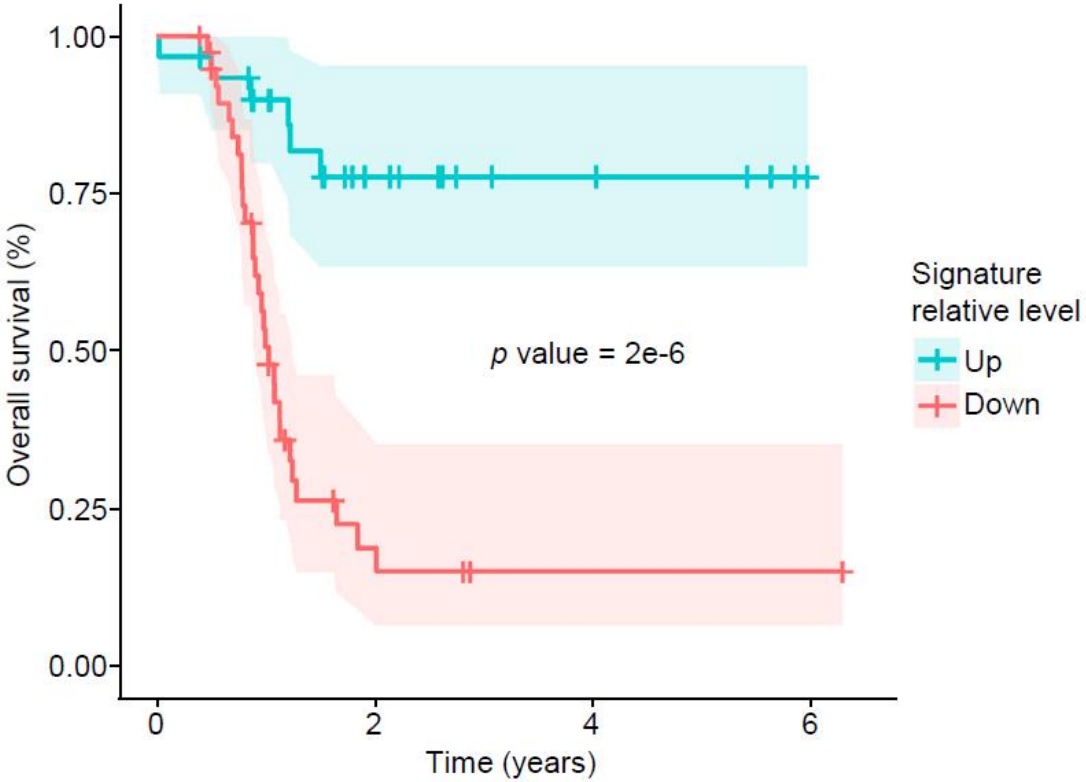
Graphical Abstract



Authors
Hui Zhang, Tao Liu, Zhen Zhang, ..., Daniel W. Chan, Karin D. Rodland, the CPTAC Investigators

Correspondence
dchan@jhmi.edu (D.W.C.), karin.rodland@pnnl.gov (K.D.R.)

In Brief
Layering proteomic and genomic data from ovarian tumors provides insights into how signaling pathways correspond to specific genome rearrangements and points to the benefit of using protein signatures for assessing prognosis and treatment stratification.



Overall Survival Stratified by CNA-Derived Signatures

Zhang et al., 2016, Cell 166, 755–765

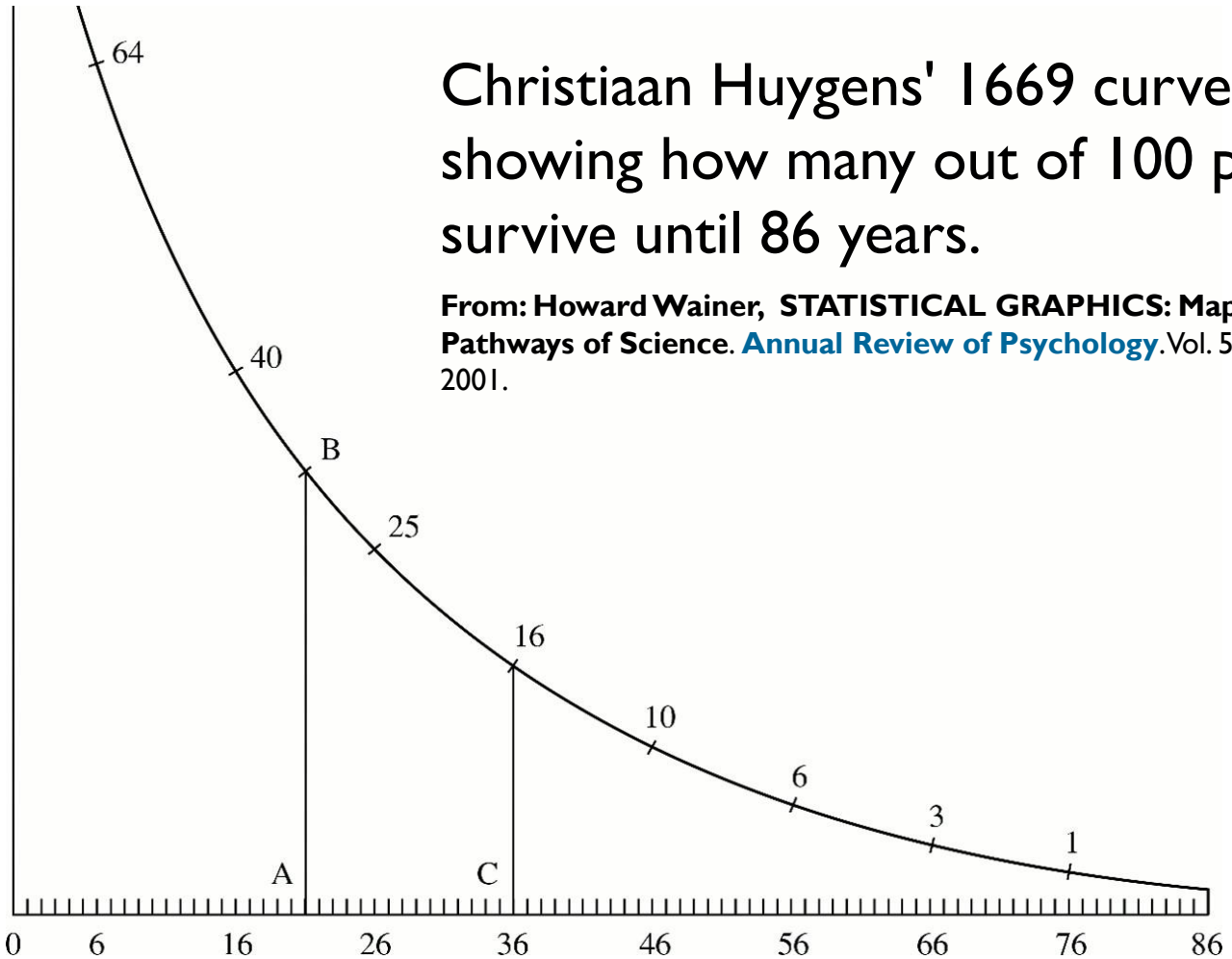
Overview of common statistical tests

Outcome Variable	Are the observations correlated?		Assumptions
	independent	correlated	
Continuous (e.g. blood pressure, age, pain score)	Ttest ANOVA Linear correlation Linear regression	Paired ttest Repeated-measures ANOVA Mixed models/GEE modeling	Outcome is normally distributed (important for small samples). Outcome and predictor have a linear relationship.
Binary or categorical (e.g. breast cancer yes/no)	Chi-square test Relative risks Logistic regression	McNemar's test Conditional logistic regression GEE modeling	Chi-square test assumes sufficient numbers in each cell (>=5)
Time-to-event (e.g. time-to-death, time-to-fracture)	Kaplan-Meier statistics Cox regression	n/a	Cox regression assumes proportional hazards between groups

Topics

- What is survival analysis?
- Terminology and data structure.
- Survival/hazard functions.
- Kaplan-Meier methods (Estimation of survival curve).
- Log-rank test (Comparison of survival curve)

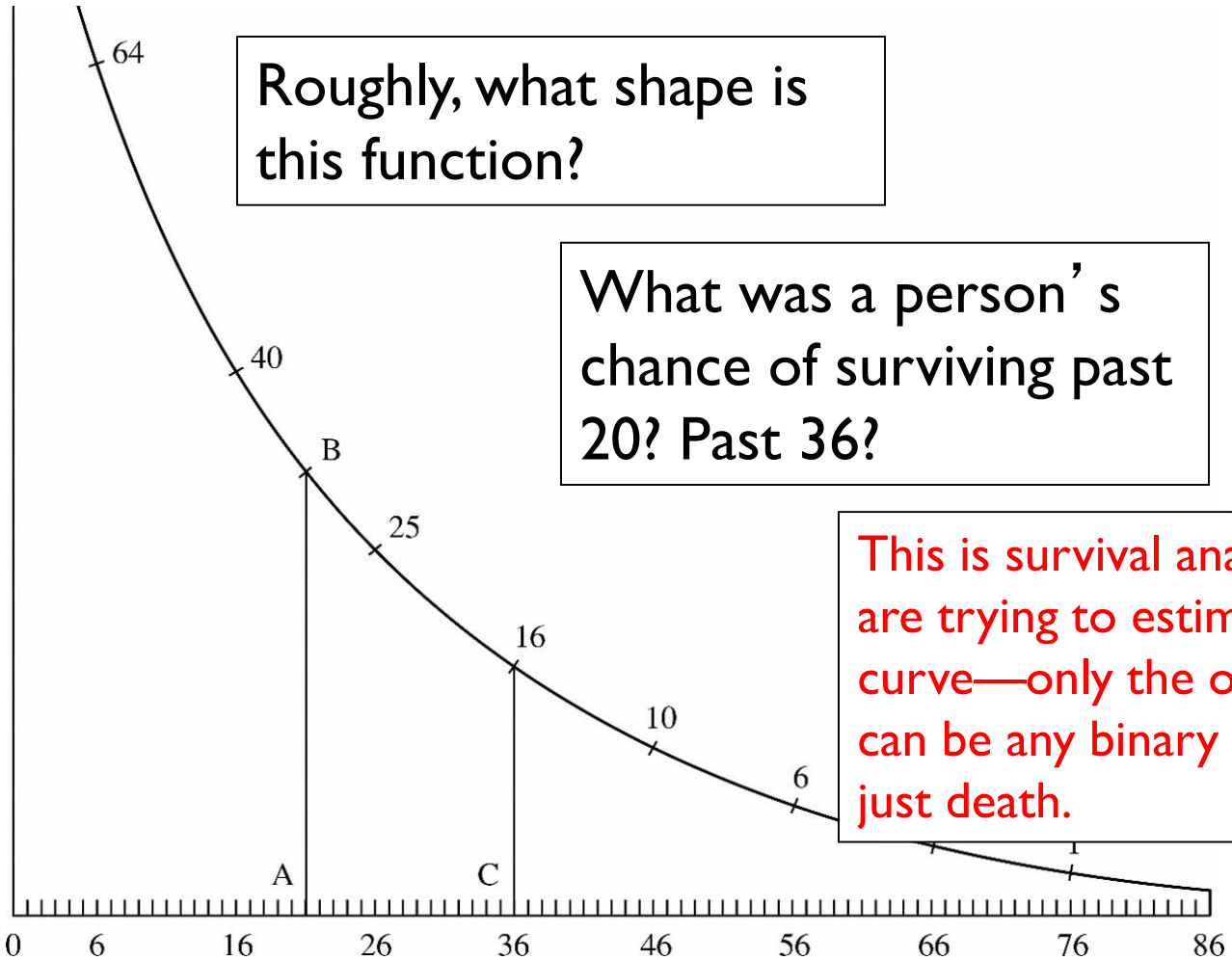
Early example of survival analysis, 1669



Christiaan Huygens' 1669 curve showing how many out of 100 people survive until 86 years.

From: Howard Wainer, STATISTICAL GRAPHICS: Mapping the Pathways of Science. Annual Review of Psychology. Vol. 52: 305-335, 2001.

Early example of survival analysis



Roughly, what shape is this function?

What was a person's chance of surviving past 20? Past 36?

This is survival analysis! We are trying to estimate this curve—only the outcome can be any binary event, not just death.

What is survival analysis?

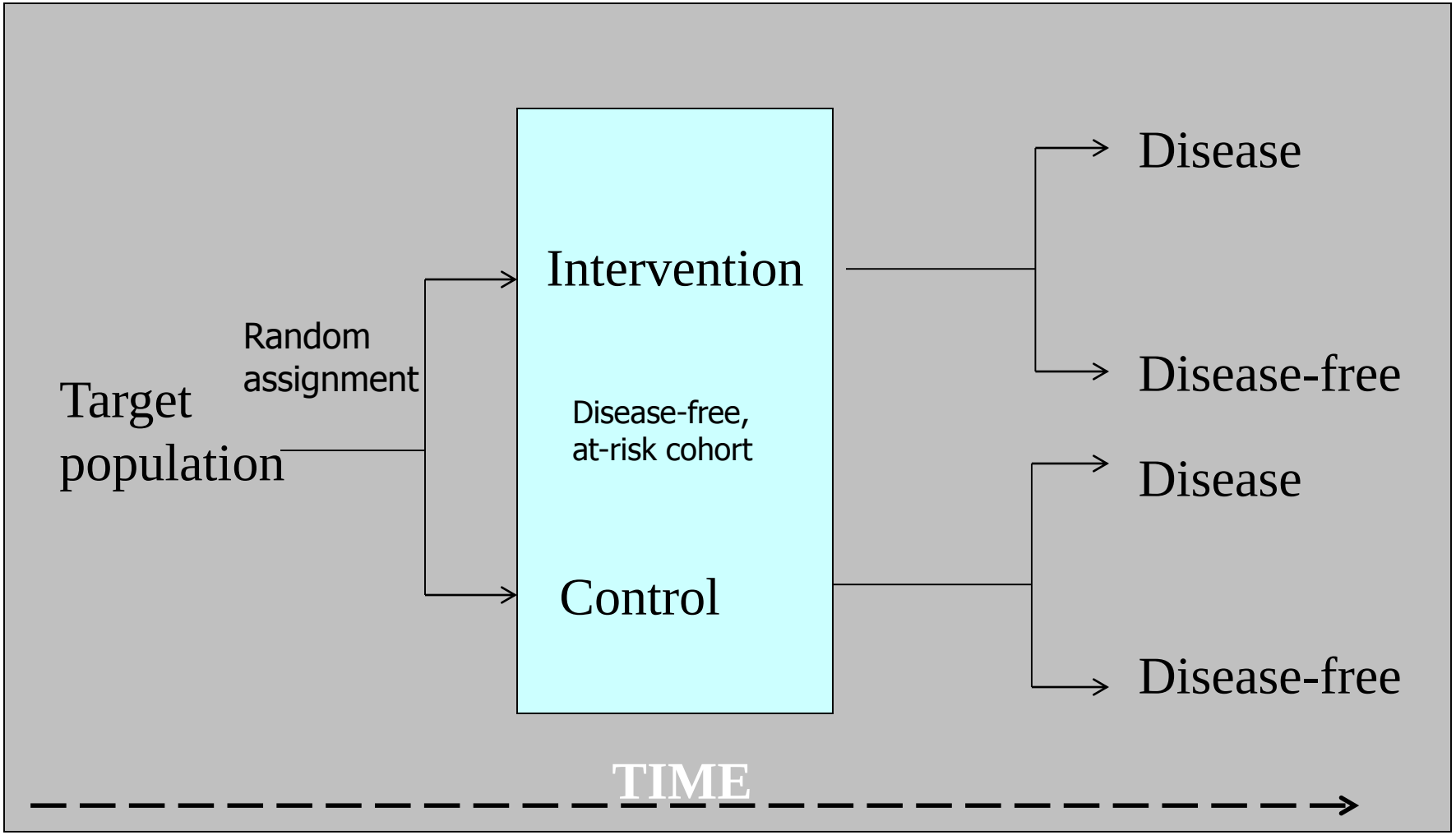
- Statistical methods for analyzing longitudinal data (纵向数据) on the occurrence of events.

* A dataset is **longitudinal** if it tracks the same type of information on the same subjects at multiple points in time

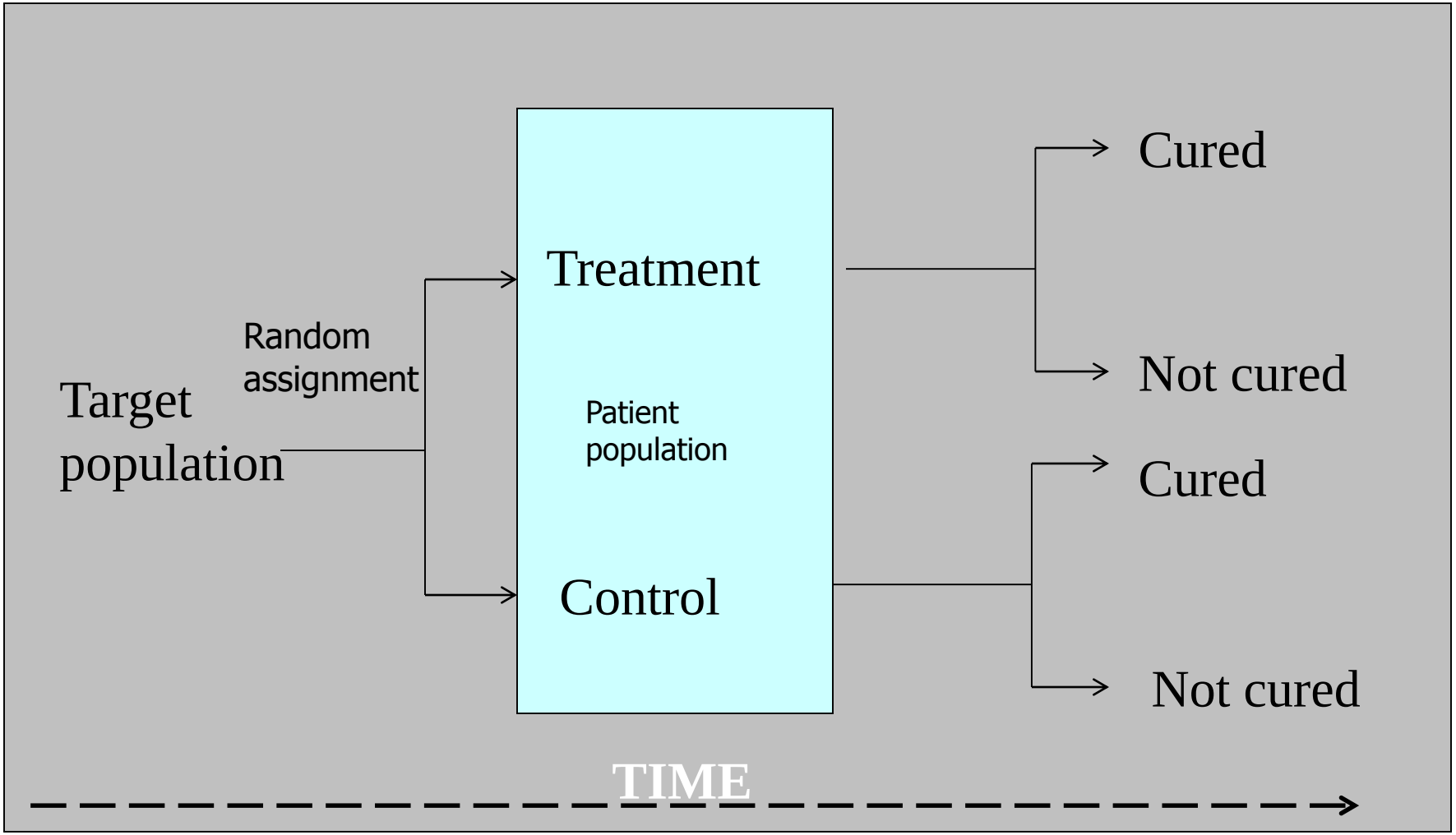
- Events may include death, injury, onset of illness, recovery from illness (binary variables) or transition above or below the clinical threshold of a meaningful continuous variable (e.g. CD4 counts->HIV).

- Accommodates data from randomized clinical trial or cohort study design（队列研究）.

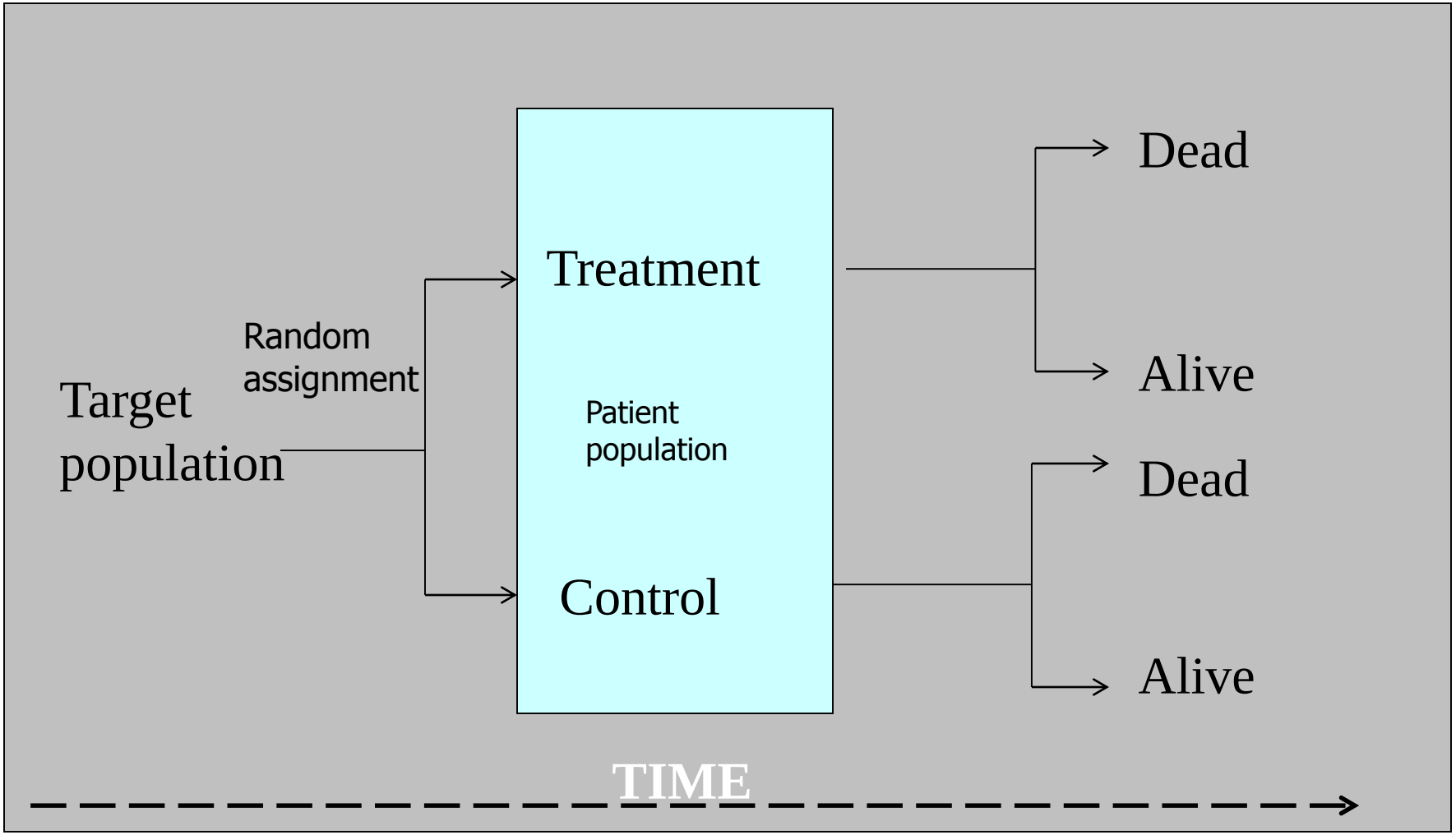
Randomized Clinical Trial (RCT) 随机临床试验



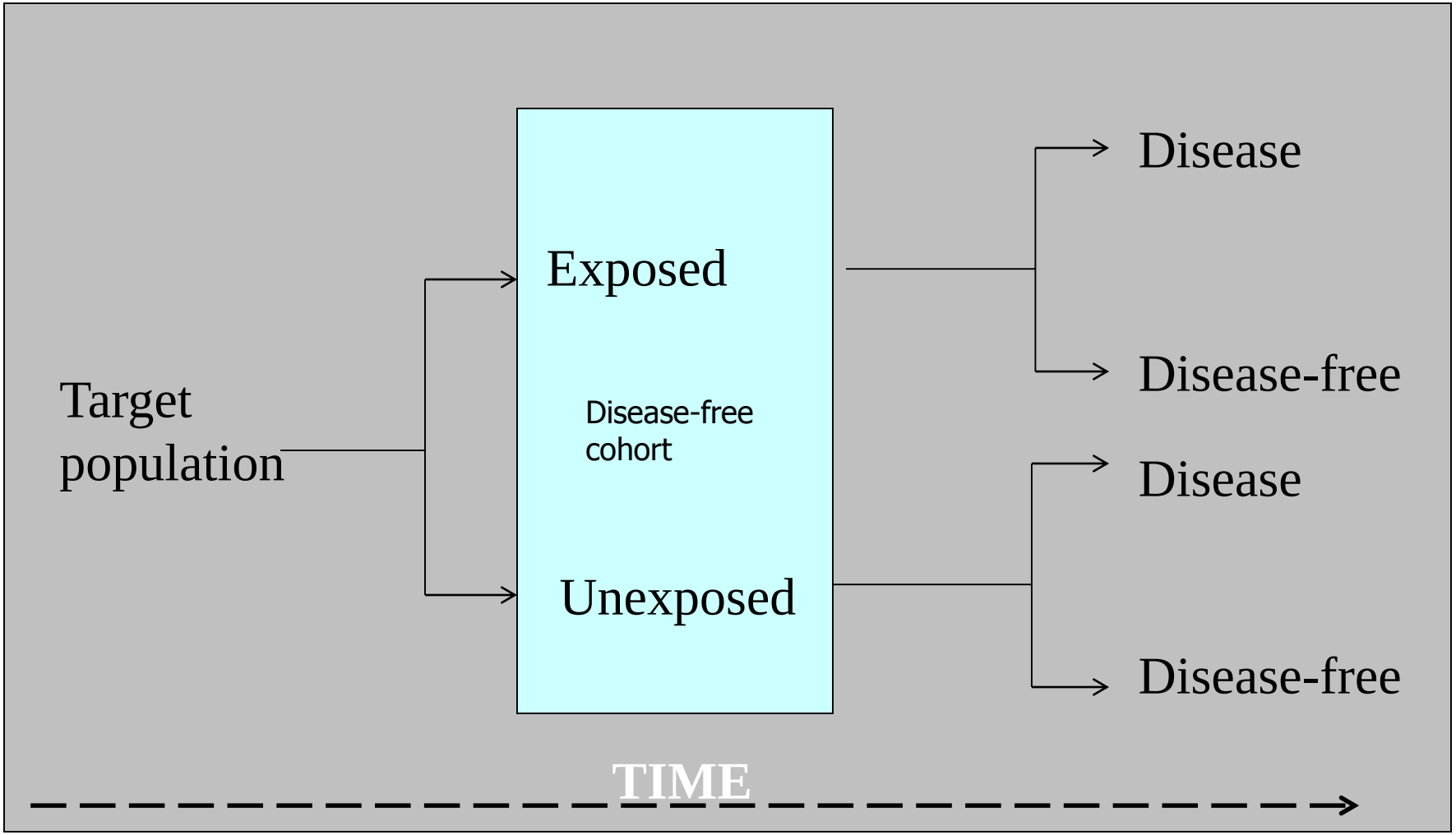
Randomized Clinical Trial (RCT) 随机临床试验



Randomized Clinical Trial (RCT) 随机临床试验



Cohort study (队列研究)



Some concepts

Risk=Cumulative incidence

=number of new cases of disease in period/number initially disease-free

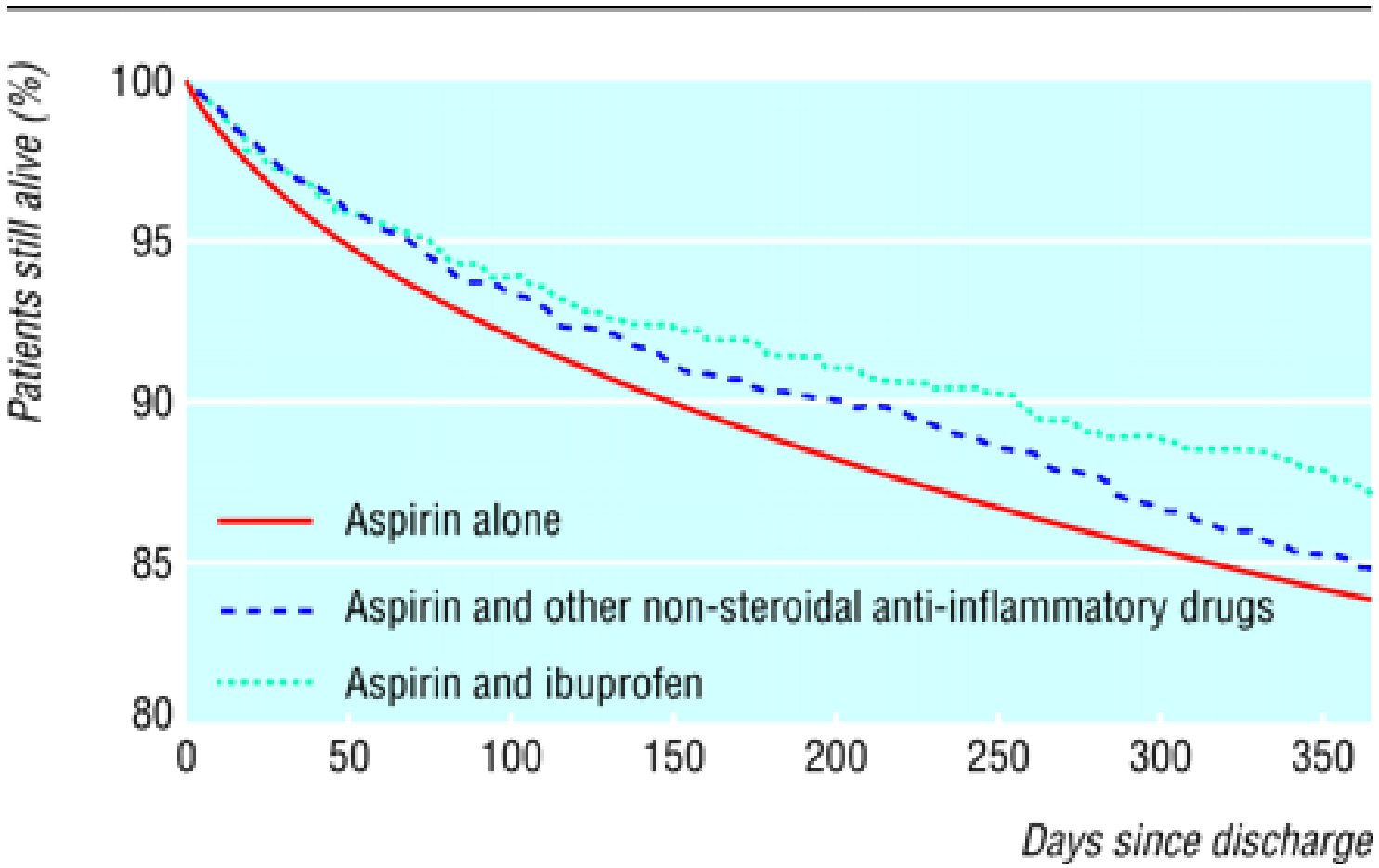
Mortality probability r(t) (死亡概率)

A measure of the number of deaths (in general, or due to a specific cause) in a particular population, scaled to the size of that population after a defined period time

Survivor probability s(t)（生存概率）=1-r(t)

The probability of survival is the probability that a person alive today will still be alive after a defined period of time.

Cohort study: Aspirin, ibuprofen, and mortality after myocardial infarction（心肌梗塞）



Curits et al. BMJ 2003;327:1322-1323.

Objectives of survival analysis

- Estimate time-to-event for a group of individuals, such as time until second heart-attack for a group of MI（heart attack,*心肌梗塞*）patients .

- To compare time-to-event between two or more groups, such as treated vs. placebo MI patients in a randomized controlled trial.

Why use survival analysis?

1. Why not compare mean time-to-event between your groups using a t-test or linear regression?

2. Why not compare proportion of events in your groups using risk/odds ratios or logistic regression?

Why use survival analysis?

1. Why not compare mean time-to-event between your groups using a t-test or linear regression?

   -- ignores censoring

2. Why not compare proportion of events in your groups using risk/odds ratios?

   --ignores time

Survival Analysis: Terms

- Time-to-event: The time from entry into a study until a subject has a particular outcome

- Censor (终检): Subjects are said to be censored if they are lost to follow up or drop out of the study, or if the study ends before they die or have an outcome of interest. They are counted as alive or disease-free for the time they were enrolled in the study.

Right Censoring (T>t)

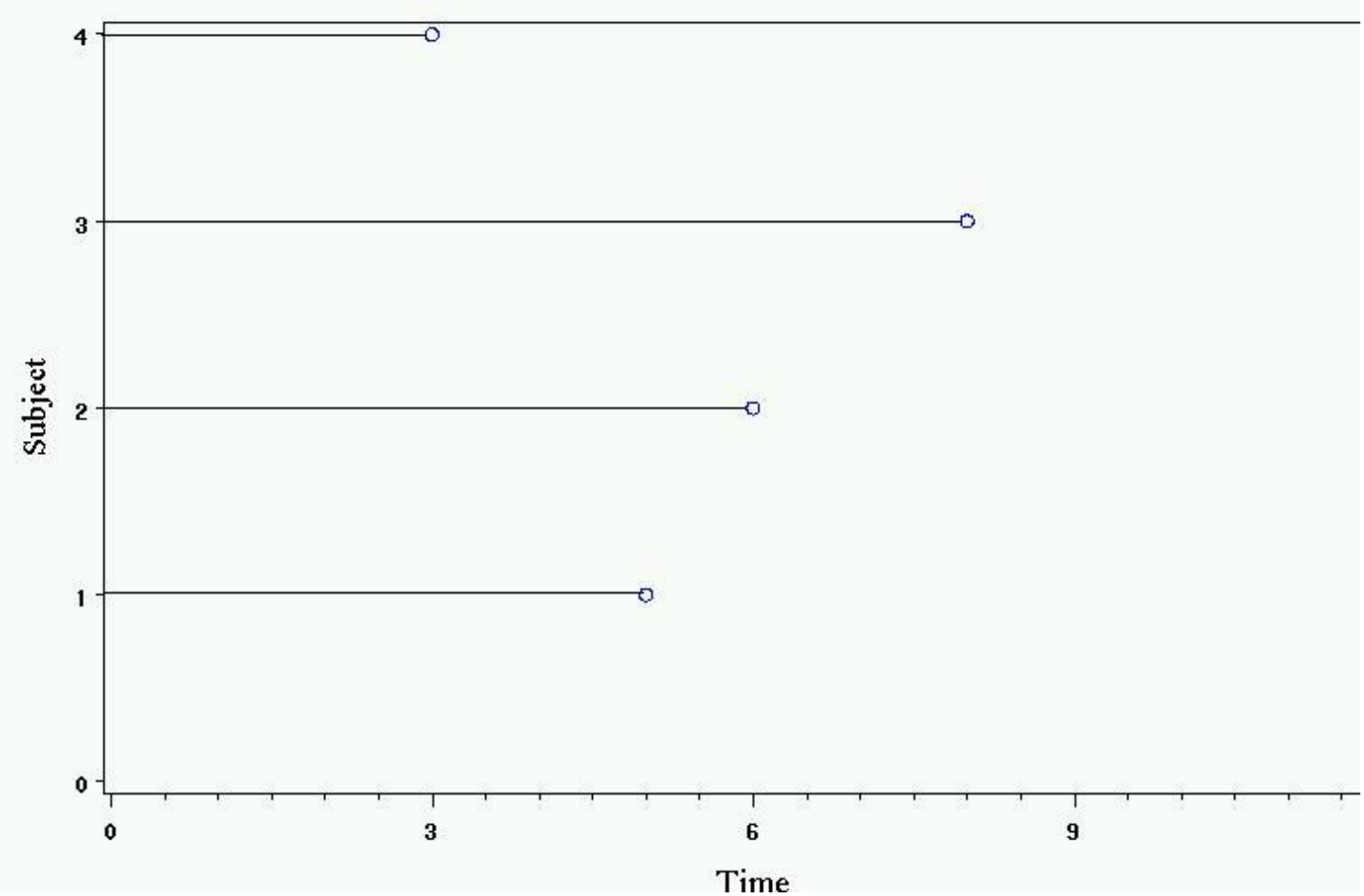
<u>Common examples</u>

- Termination of the study
- Death due to a cause that is not the event of interest
- Loss to follow-up

We know that subject survived at least to time *t*.

Count every subject's time since their baseline data collection.

Right-censoring!



Probability density function: f(t)

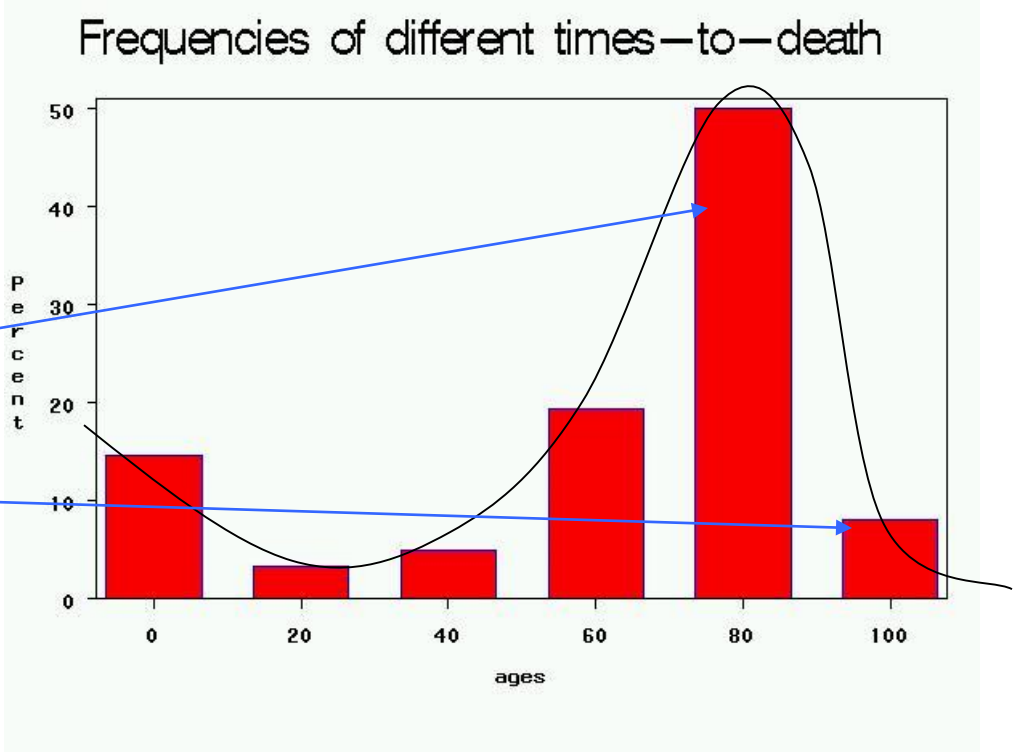
Survival Distribution

In the case of human longevity, T_i is unlikely to follow a normal distribution, because the probability of death is not highest in the middle ages, but at the beginning and end of life.

Hypothetical data:

People have a high chance of dying in their 70's and 80's;

BUT they have a smaller chance of dying in their 90's and 100's, because few people make it long enough to die at these ages.



Survival function

The goal of survival analysis is to estimate and compare survival experiences of different groups.

Survivor function S(t)（生存函数）, illustated by the survival curve. This is the probality that an individual will survive (i.e. has not experienced the event of interest) up to and including time t

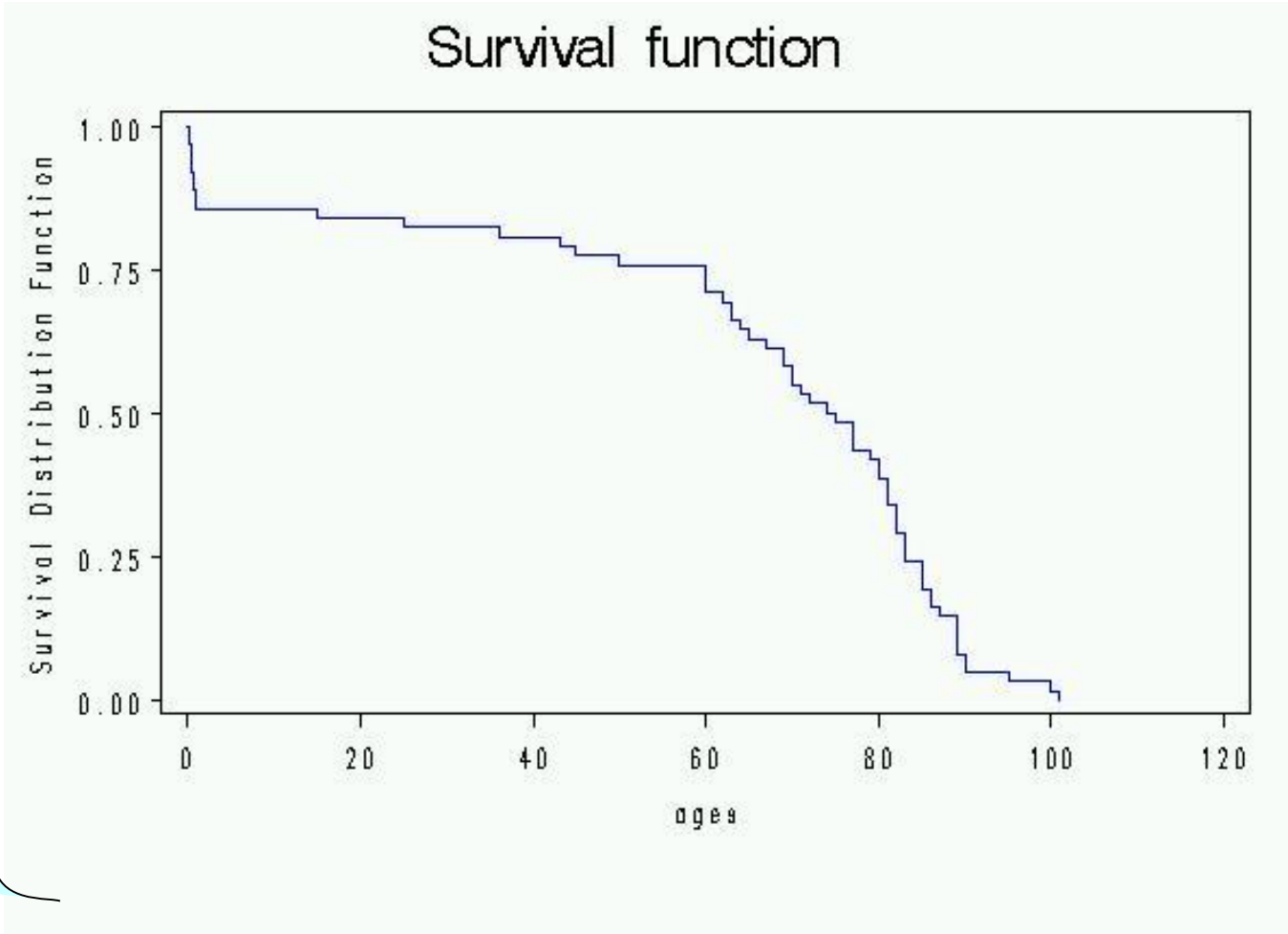
$$S(t) = 1 - P(T £ t)$$

Example: If t=100 years, S(t=100) = probability of surviving beyond 100 years.

（个体生存时间大于100年d概率）

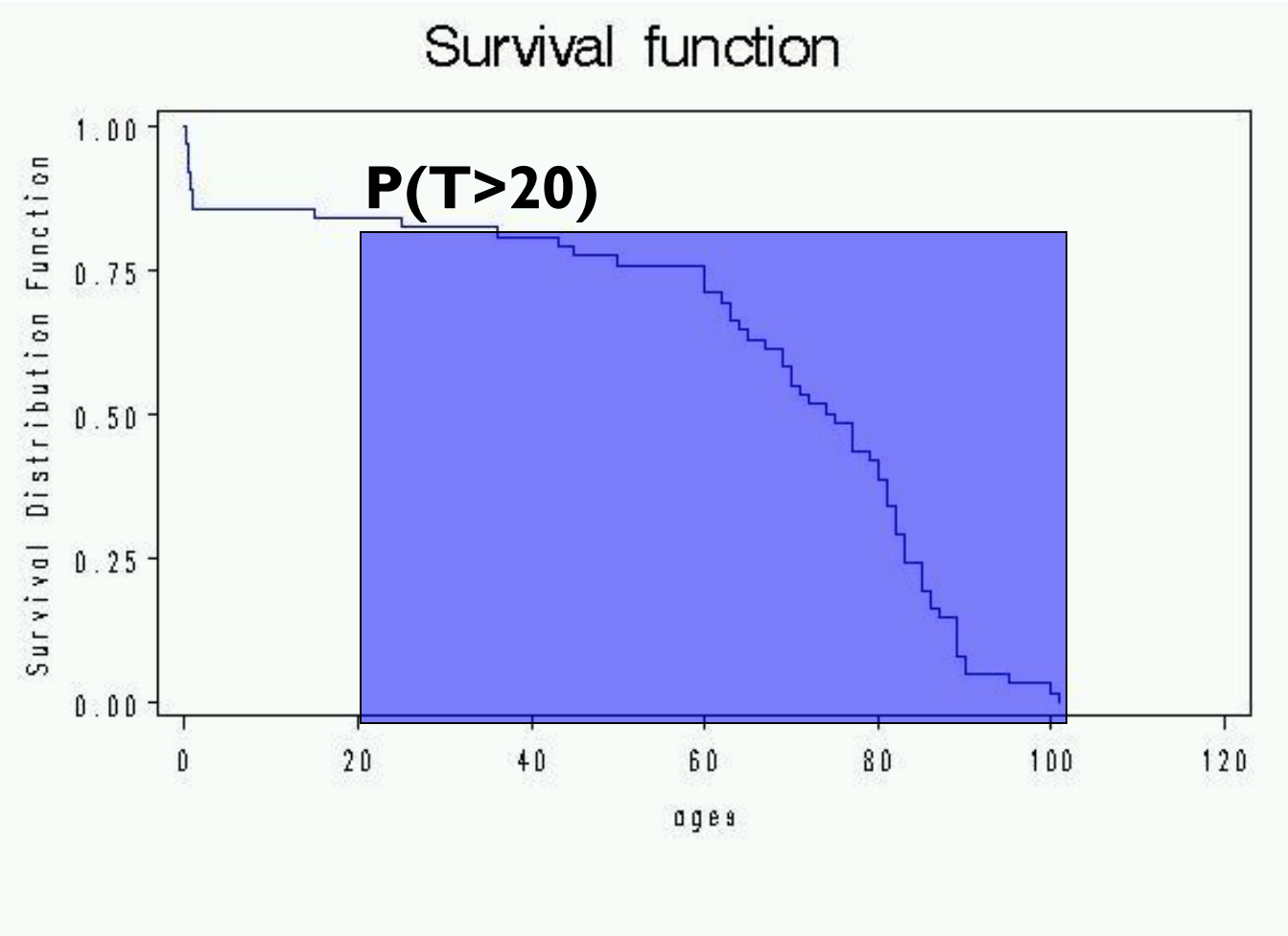
Cumulative survival

Same hypothetical data, plotted as cumulative distribution rather than density:

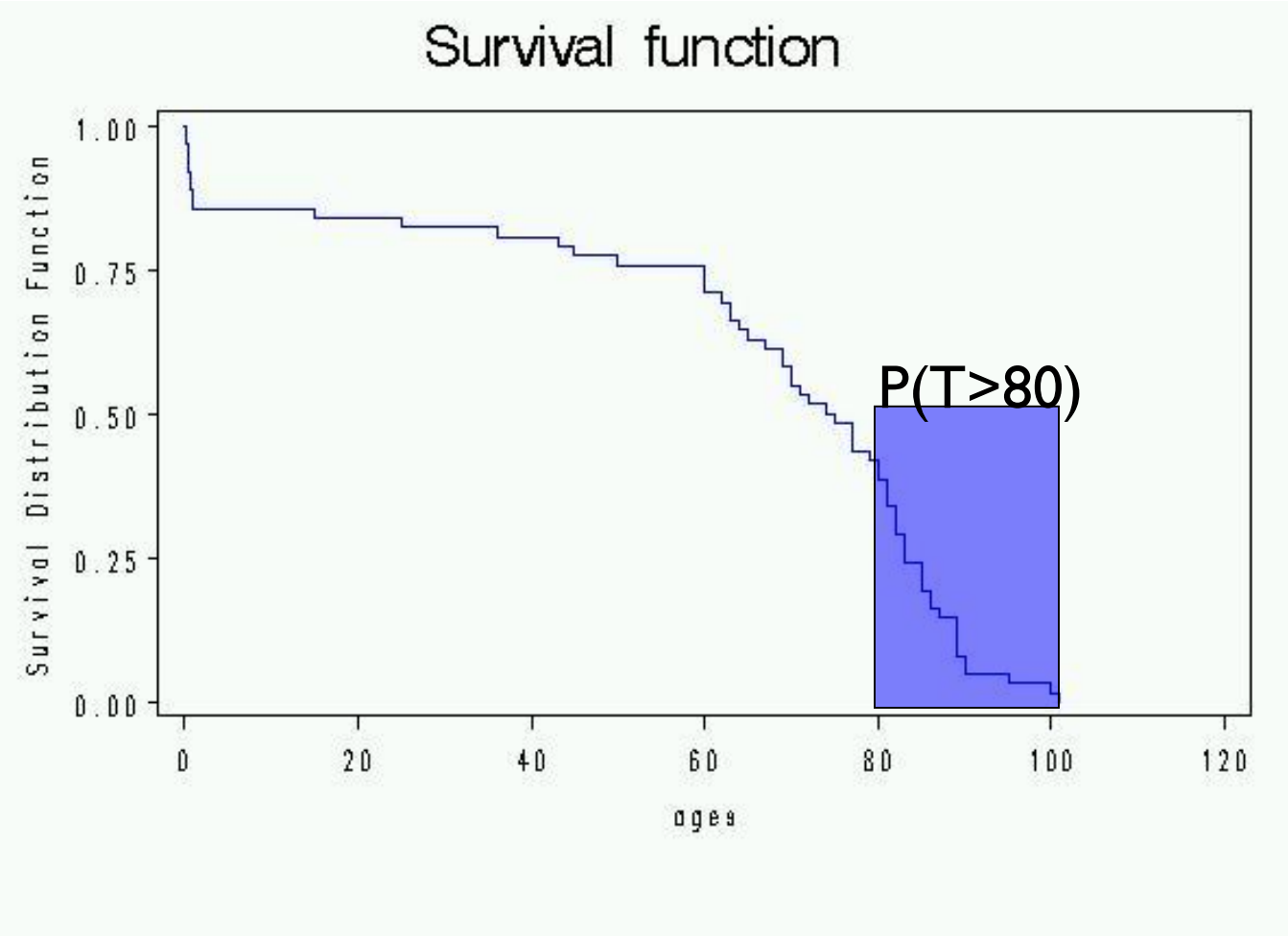


Recall pdf:

Cumulative survival



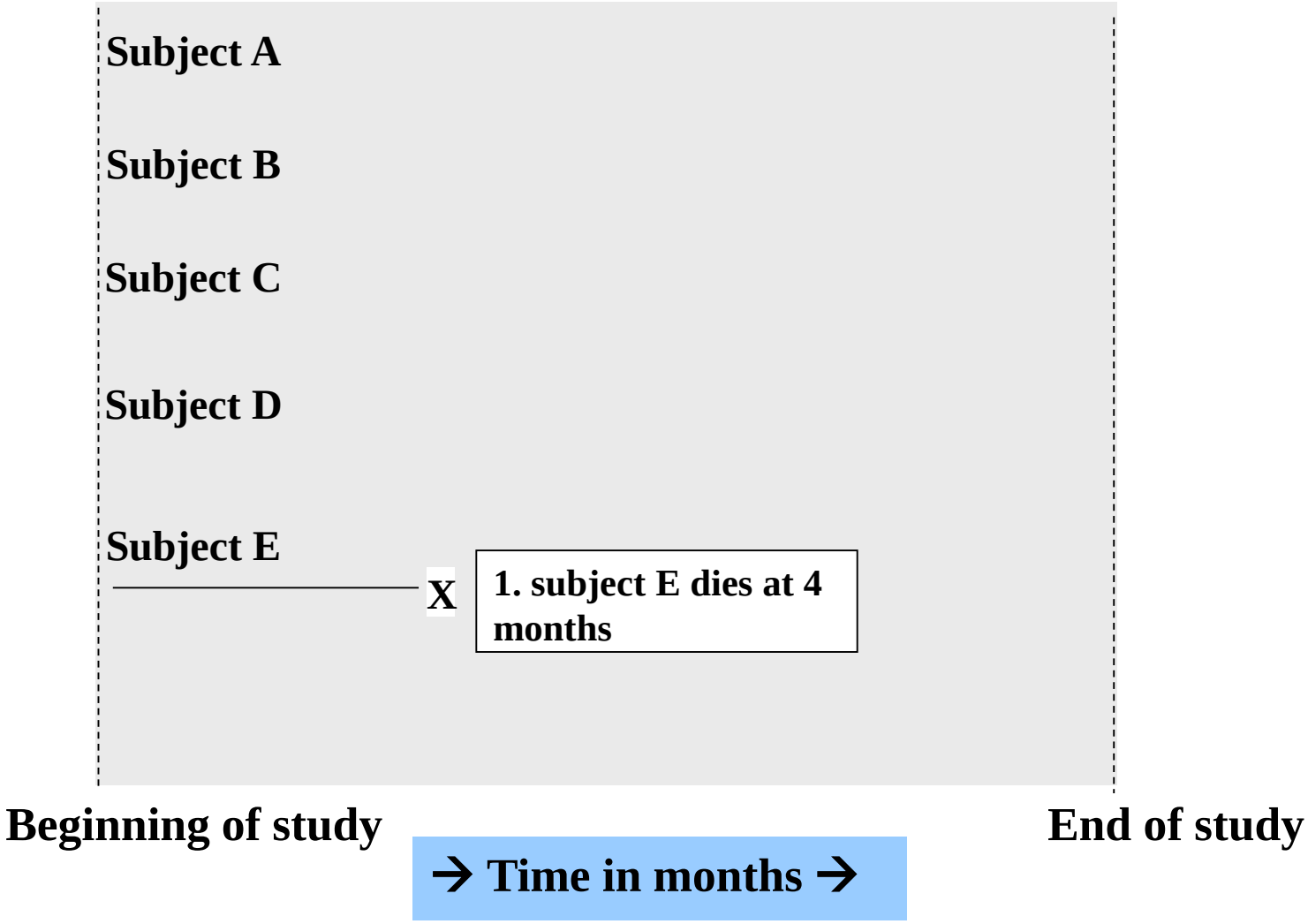
Cumulative survival



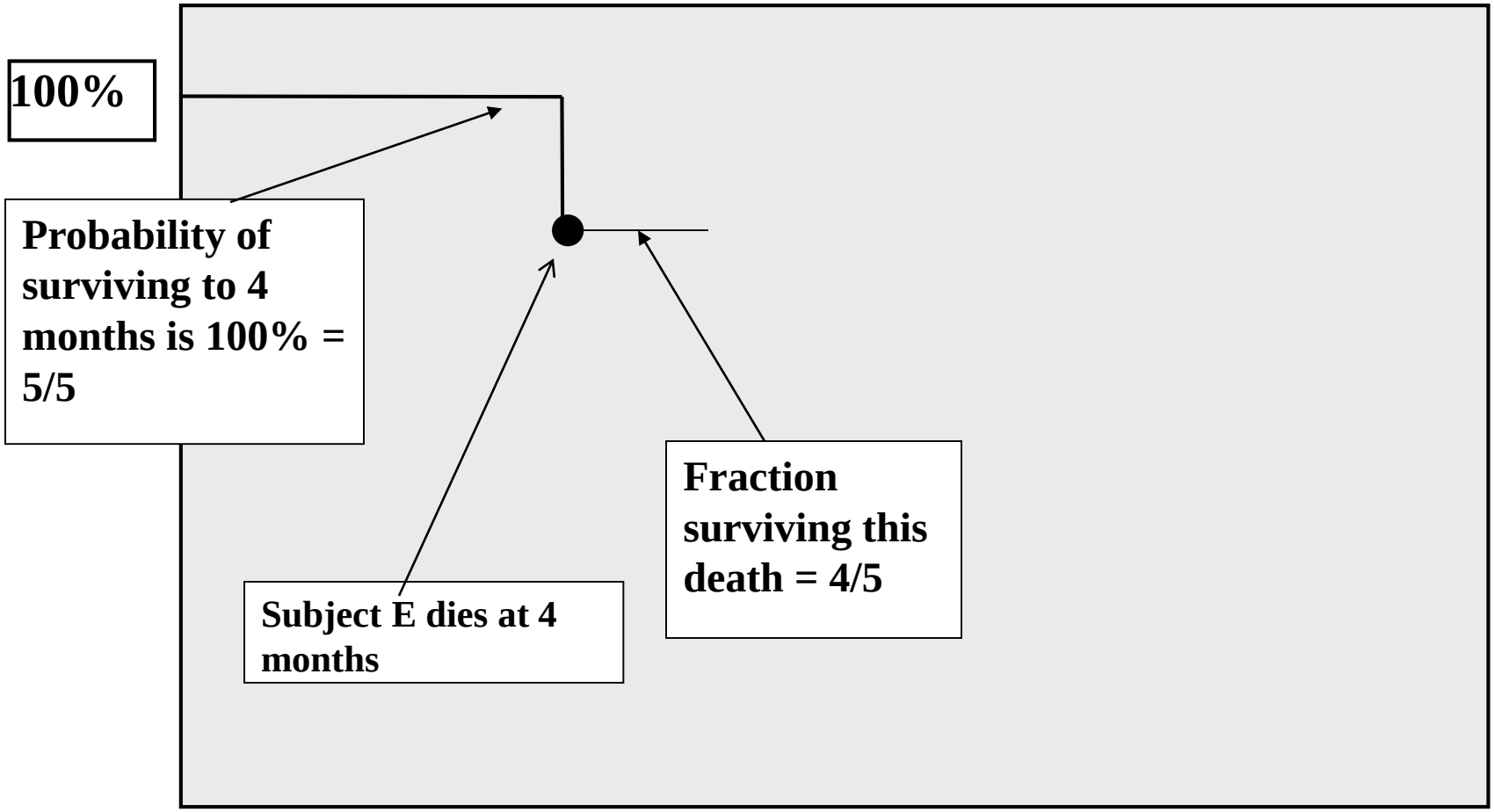
Kaplan-Meier（KM）method

- Non-parametric estimate of the survival function.
- Commonly used to describe survivorship of study population/s.
- Commonly used to compare two study populations.
- Intuitive graphical presentation.

Survival Data (right-censored)

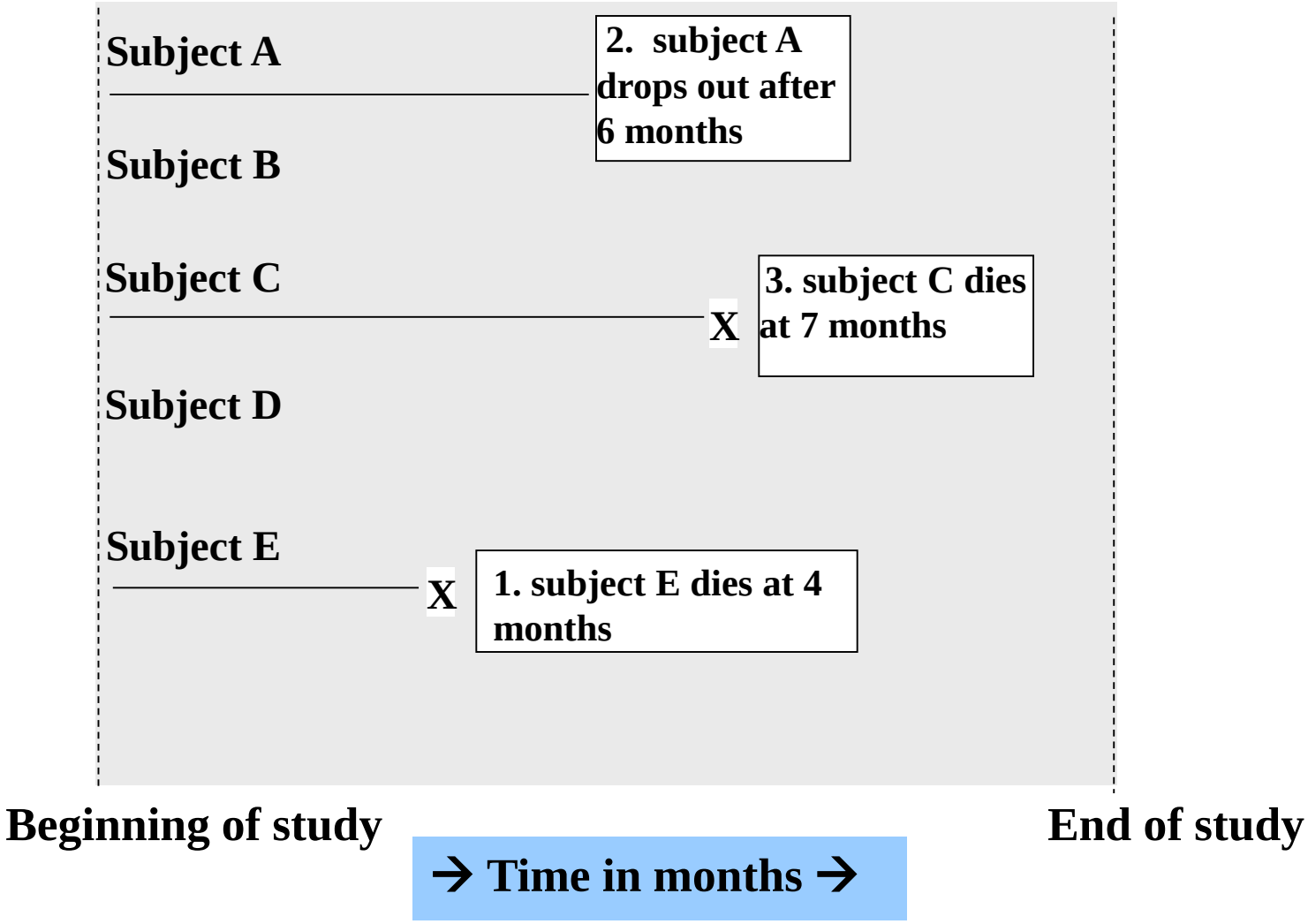


Corresponding Kaplan-Meier Curve

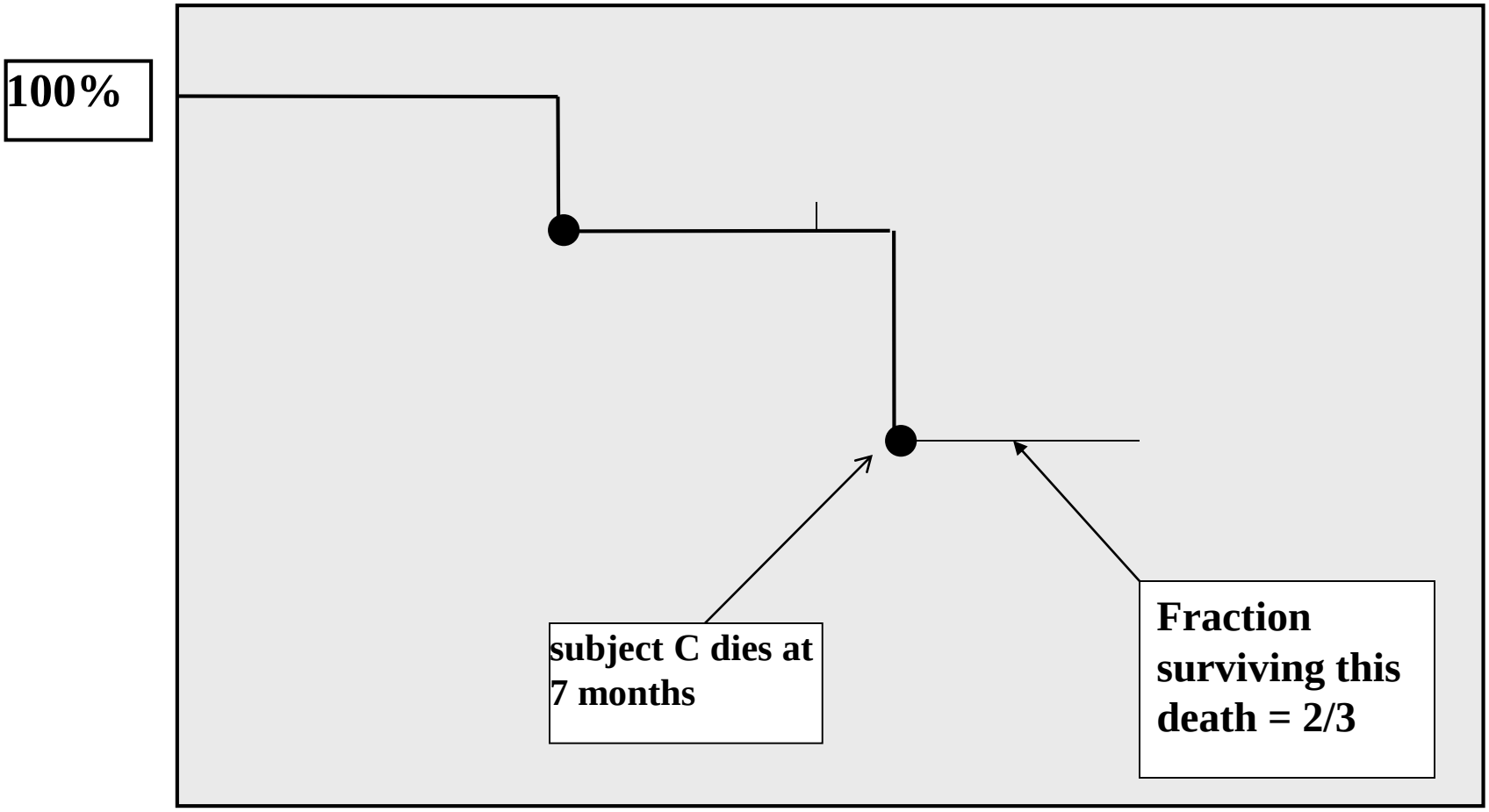


→ Time in months →

Survival Data

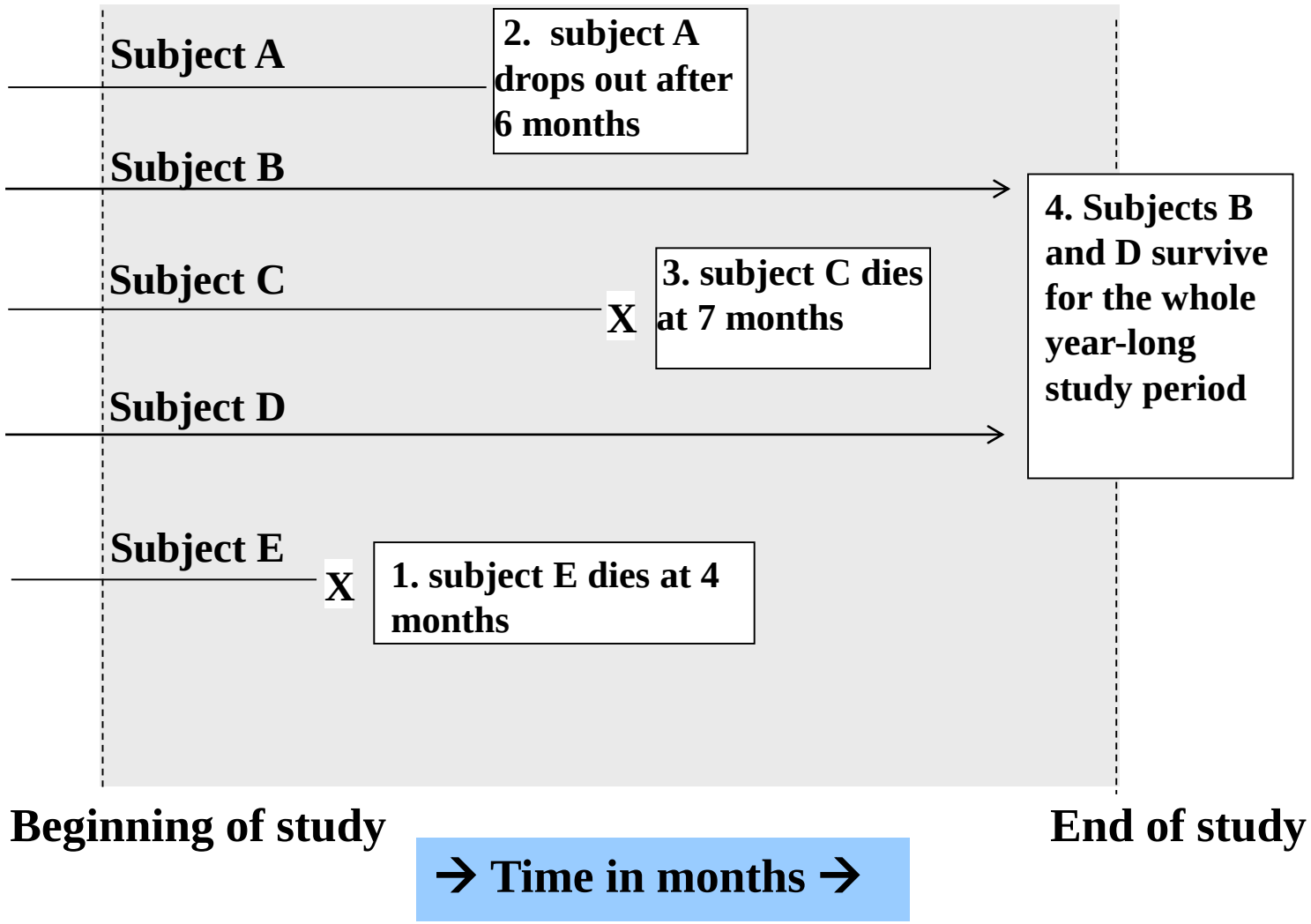


Corresponding Kaplan-Meier Curve

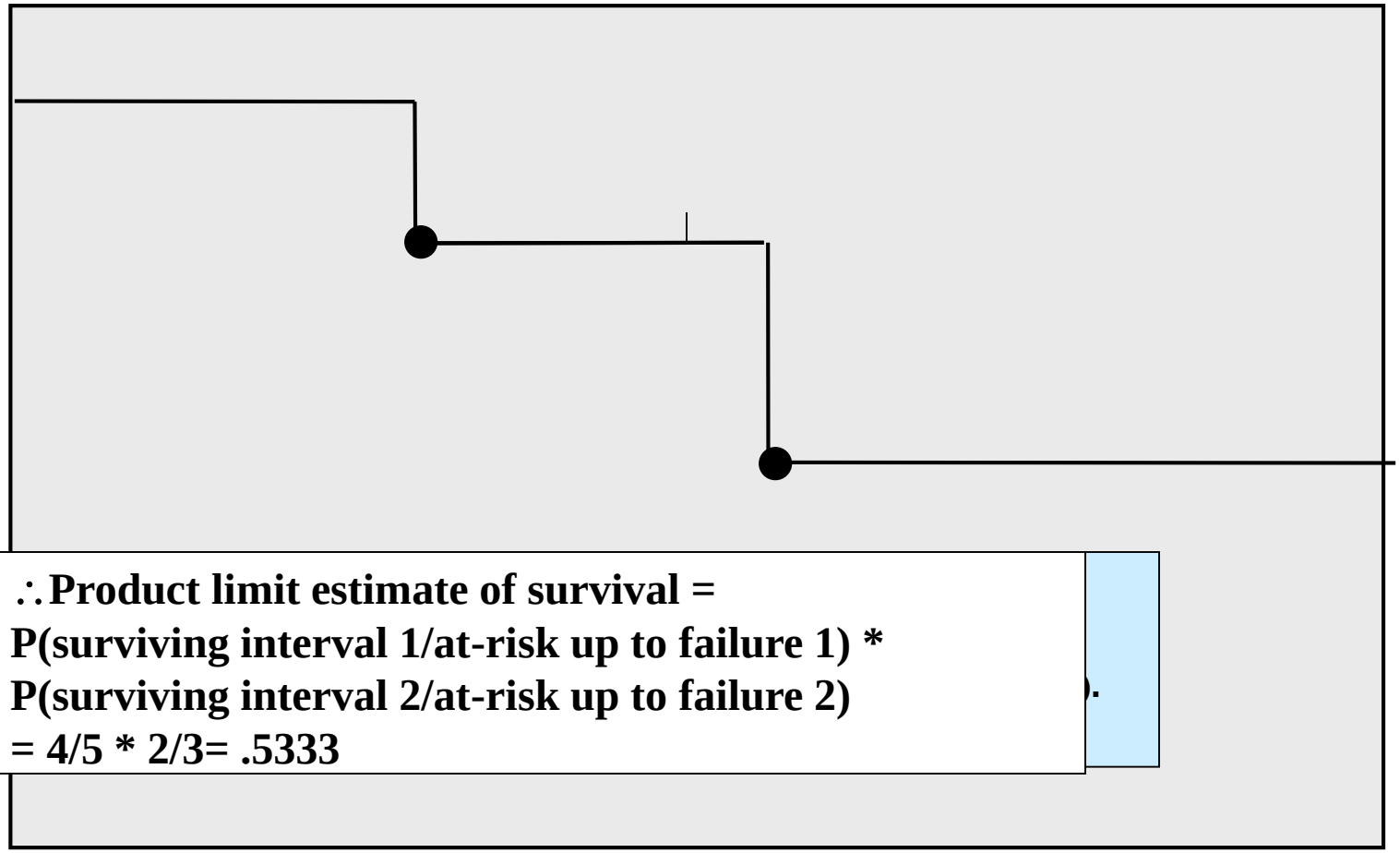


→ Time in months →

Survival Data



Corresponding Kaplan-Meier Curve



100%

Rule from prob

P(A&B)=P(A)*

In survival ana

P(surviving int

∴Product limit estimate of survival =
P(surviving interval 1/at-risk up to failure 1) *
P(surviving interval 2/at-risk up to failure 2)
= 4/5 * 2/3= .5333

乘积极限法

→ Time in months →

The product limit estimate

- The probability of surviving in the entire year, taking into account censoring
  = (4/5) (2/3) = 53%

- NOTE: > 40% (2/5) because the one drop-out survived at least a portion of the year.

- AND <60% (3/5) because we don't know if the one drop-out would have survived until the end of the year.

Kaplan-Meier estimate of the survival curve

Survival probability at time t is

$$s_t = 1 - r_t = \frac{n_t - d_t}{n_t}$$

Product-limit formula
乘积极限法

Survival function (cumulative survival probability)

$$S(t_1) = 1 \times s_{t1} = s_{t1}$$

$$S(t_2) = S(t_1) \times s_{t2} = s_{t1} \times s_{t2}$$

In general $$S(t_j) = S(t_{(j-1)}) \times s_{tj} = s_{t1} \times s_{t2} \times \cdots \times s_{tj}$$

Example - Navelbine/Taxol vs Leukemia

- Mice given P388 murine leukemia assigned at random to one of two regimens of therapy
  - Regimen A – Navelbine（双酒石酸盐）+ Taxol (紫杉醇) Concurrently
  - Regimen B - Navelbine + Taxol 1-hour later

- Under regimen A, 9 of n_A=49 mice died on days: 6,8,22,32,32,35,41,46, and 54. Remainder > 60 days
- Under regimen B, 9 of n_B=15 mice died on days: 8,10,27,31,34,35,39,47, and 57. Remainder > 60 days

Source: Knick, et al (1995)

Example - Navelbine/Taxol vs Leukemia

Regimen A

i	$t_{(i)}$	n_i	d_i	λ_i	$S(t_{(i)})$
1	6	49			
2	8	48			
3	22	47			
4	32	46			
5	35	44			
6	41	43			
7	46	42			
8	54	41			

Regimen B

i	$t_{(i)}$	n_i	d_i	λ_i	$S(t_{(i)})$
1	8	15			
2	10	14			
3	27	13			
4	31	12			
5	34	11			
6	35	10			
7	39	9			
8	47	8			
9	57	7			

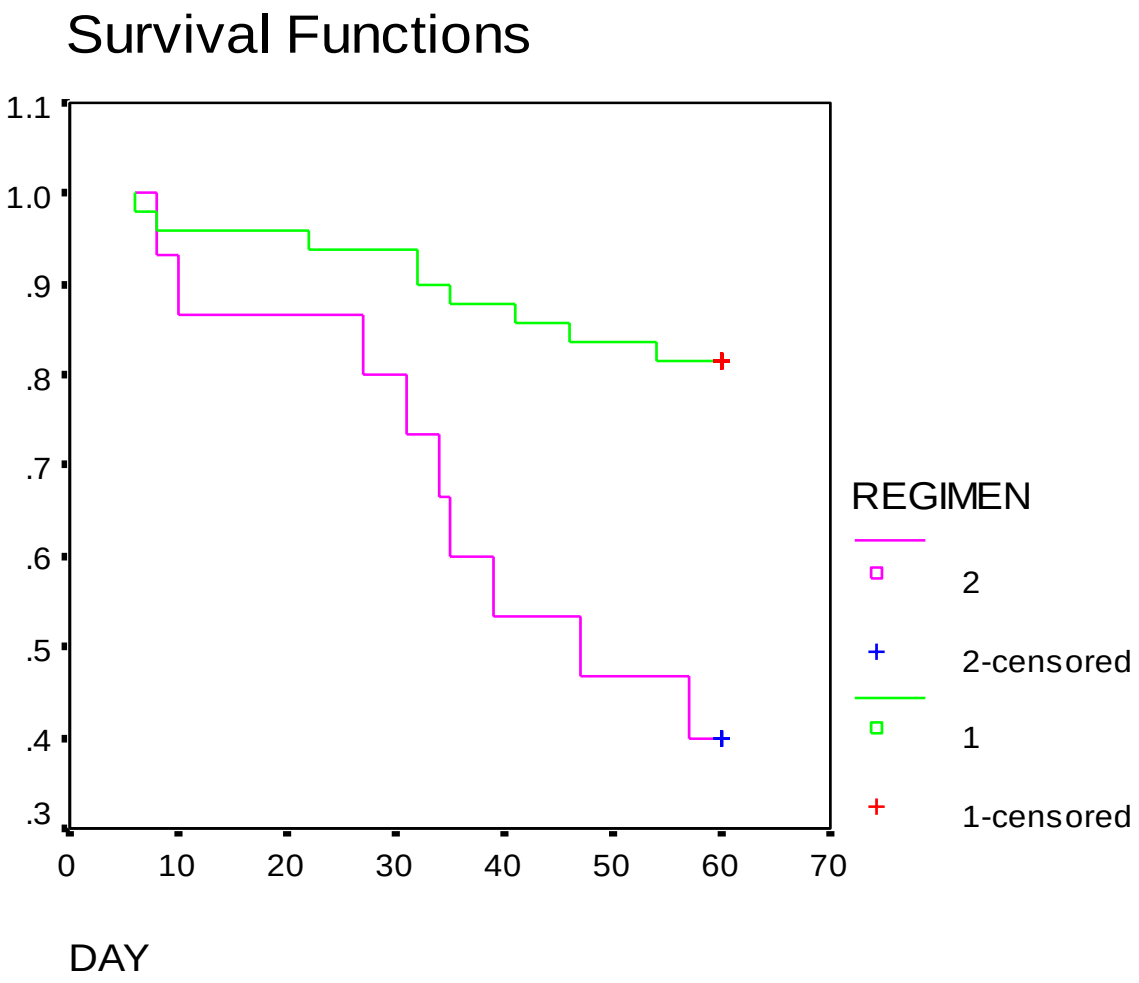
Example - Navelbine/Taxol vs Leukemia

Regimen A						Regimen B					
i	$t_{(i)}$	n_i	d_i	λ_i	$S(t_{(i)})$	i	$t_{(i)}$	n_i	d_i	λ_i	$S(t_{(i)})$
1	6	49	1	.020	.980	1	8	15	1	.067	.933
2	8	48	1	.021	.959	2	10	14	1	.071	.867
3	22	47	1	.021	.939	3	27	13	1	.077	.800
4	32	46	2	.043	.899	4	31	12	1	.083	.733
5	35	44	1	.023	.878	5	34	11	1	.091	.667
6	41	43	1	.023	.858	6	35	10	1	.100	.600
7	46	42	1	.024	.837	7	39	9	1	.111	.533
8	54	41	1	.024	.817	8	47	8	1	.125	.467
						9	57	7	1	.143	.400

$$\hat{\lambda}_1^A = \frac{1}{49} = .020 \qquad \hat{S}^A(6) = 1 - .020 = .980$$

$$\hat{\lambda}_2^A = \frac{1}{48} = .021 \qquad \hat{S}^A(8) = .980(1 - .021) = .959$$

Example - Navelbine/Taxol vs Leukemia



Kaplan-Meier: another example

Researchers randomized 44 patients with chronic active hepatitis（慢性肝炎） were to receive prednisolone(脱氢皮质醇, 一种糖皮质激素) or no treatment (control), then compared survival curves.

Example from: BMJ 1998;317:468-469 (15 August)

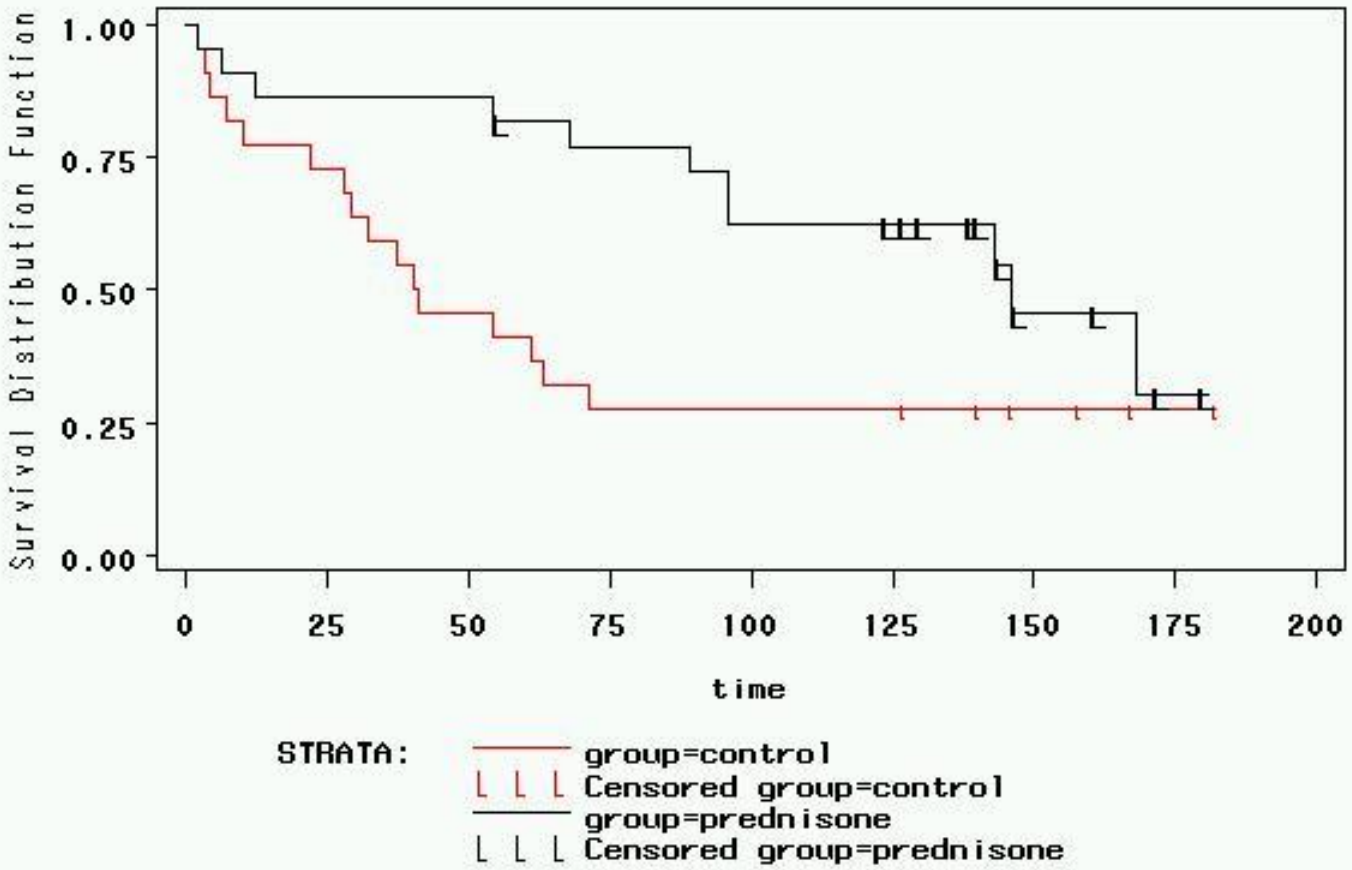
Survival times (months) of 44 patients with chronic active hepatitis randomised to receive prednisolone or no treatment.

Prednisolone (n=22)	Control (n=22)
2	2
6	3
12	4
54	7
56 *	10
68	22
89	28
96	29
96	32
125*	37
128*	40
131*	41
140*	54
141*	61
143	63
145*	71
146	127*
148*	140*
162*	146*
168	158*
173*	167*
181*	182*

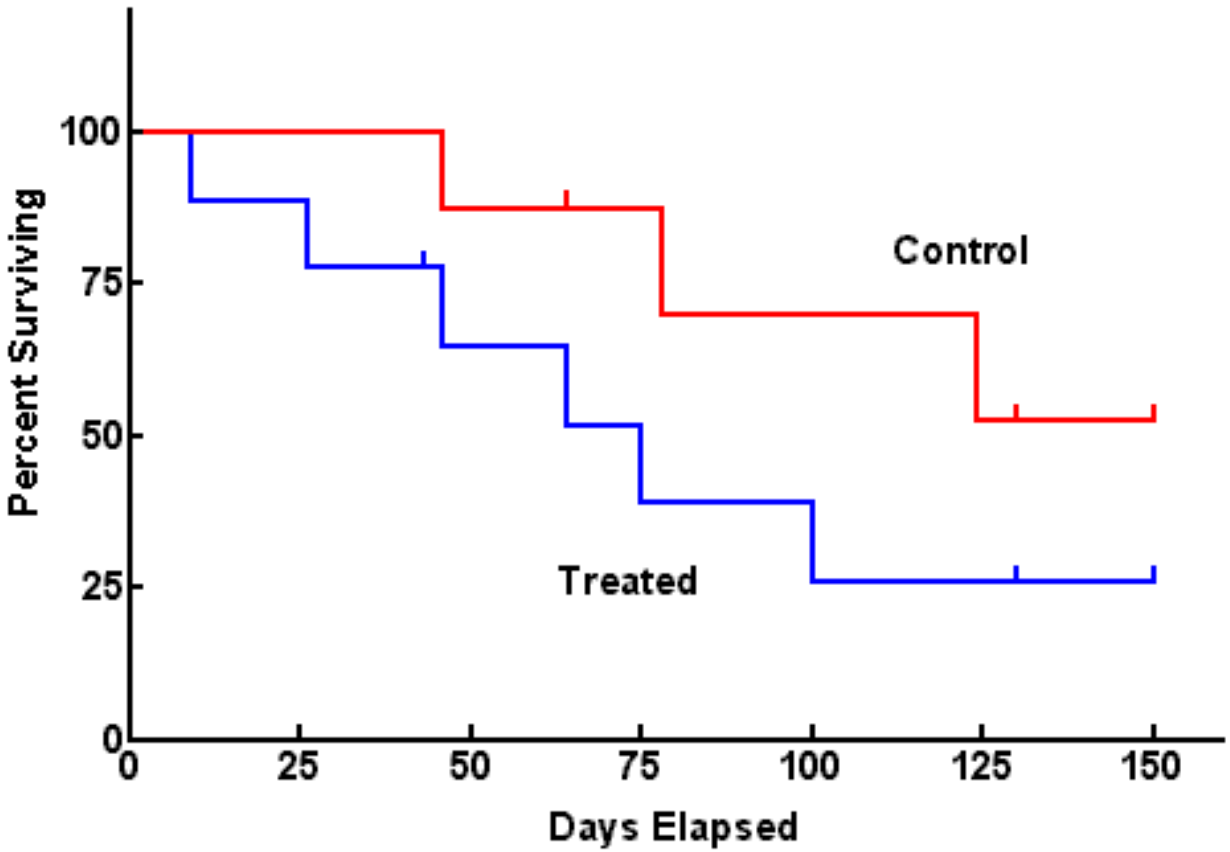
Data from: BMJ 1998;317:468-469 (15 August) *=censored (终检/截尾)

Kaplan-Meier: example

Are these two curves different?



Comparing two survival curves



Use log-rank test to test the null hypothesis of no difference between survival functions of the two groups

Example in last class

Regimen A

i	$t_{(i)}$	n_i	d_i	λ_i	$S(t_{(i)})$
1	6	49	1	.020	.980
2	8	48	1	.021	.959
3	22	47	1	.021	.939
4	32	46	2	.043	.899
5	35	44	1	.023	.878
6	41	43	1	.023	.858
7	46	42	1	.024	.837
8	54	41	1	.024	.817

Regimen B

i	$t_{(i)}$	n_i	d_i	λ_i	$S(t_{(i)})$
1	8	15	1	.067	.933
2	10	14	1	.071	.867
3	27	13	1	.077	.800
4	31	12	1	.083	.733
5	34	11	1	.091	.667
6	35	10	1	.100	.600
7	39	9	1	.111	.533
8	47	8	1	.125	.467
9	57	7	1	.143	.400

$t_{(1i)}$ n_{1i} d_{1i} $t_{(0i)}$ n_{0i} d_{0i}

Log rank test

- H_0: Two Survival Functions are Identical
- H_A: Two Survival Functions Differ

$$c^2_{MC} = \frac{U^2}{V}; \quad df = 1$$, where

$$U = \sum (d_{1i} - E_{1i}), \quad and \quad E_{1i} = \frac{d_i \times n_{1i}}{n_i}$$

$$V = \sum V_i = \sum \frac{d_i \times n_{0i} \times n_{1i}}{n_i^2}$$

Example

缓解

The data: remission times (weeks) for two groups of leukemia patients

Group 1 (n=21) treatment	Group 2 (n=21) placebo
6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+. 32+, 34+, 35+	1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

+ denotes censored

	# failed	# censored	Total
Group 1	9	12	21
Group 2	21	0	21

Descriptive statistic:

$\overline{T}_1(ignoring + 's) = 17.1, \quad \overline{T}_2 = 8.6$

Remission data: n=42

# failures			# in risk set	
$t_{(j)}$	m_{1j}	m_{2j}	n_{1j}	n_{2j}
1	0	2	21	21
2	0	2	21	19
3	0	1	21	17
4	0	2	21	16
5	0	2	21	14
6	3	0	21	12
7	1	0	17	12
8	0	4	16	12
10	1	0	15	8
11	0	2	13	8
12	0	2	12	6
13	1	0	12	4
15	0	1	11	4
16	1	0	11	3
17	0	1	10	3
22	1	1	7	2
23	1	1	6	1

Expected cell counts:

$$e_{1j} = \left(\frac{n_{1j}}{n_{1j} + n_{2j}} \right) \times \left(m_{1j} + m_{2j} \right)$$

Proportion in risk set

# of failures over both groups

$$e_{2j} = \left(\frac{n_{2j}}{n_{1j} + n_{2j}} \right) \times \left(m_{1j} + m_{2j} \right)$$

Example: Remission data

EXAMPLE

Expanded Table (Remission Data)

		# failures		# in risk set		# expected		Observed–expected	
j	$t_{(j)}$	m_{1j}	m_{2j}	n_{1j}	n_{2j}	e_{1j}	e_{2j}	$m_{1j}-e_{1j}$	$m_{2j}-e_{2j}$
1	1	0	2	21	21	(21/42) × 2	(21/42) × 2	−1.00	1.00
2	2	0	2	21	19	(21/40) × 2	(19/40) × 2	−1.05	1.05
3	3	0	1	21	17	(21/38) × 1	(17/38) × 1	−0.55	0.55
4	4	0	2	21	16	(21/37) × 2	(16/37) × 2	−1.14	1.14
5	5	0	2	21	14	(21/35) × 2	(14/35) × 2	−1.20	1.20
6	6	3	0	21	12	(21/33) × 3	(12/33) × 3	1.09	−1.09
7	7	1	0	17	12	(17/29) × 1	(12/29) × 1	0.41	−0.41
8	8	0	4	16	12	(16/28) × 4	(12/28) × 4	−2.29	2.29
9	10	1	0	15	8	(15/23) × 1	(8/23) × 1	0.35	−0.35
10	11	0	2	13	8	(13/21) × 2	(8/21) × 2	−1.24	1.24
11	12	0	2	12	6	(12/18) × 2	(6/18) × 2	−1.33	1.33
12	13	1	0	12	4	(12/16) × 1	(4/16) × 1	0.25	−0.25
13	15	0	1	11	4	(11/15) × 1	(4/15) × 1	−0.73	0.73
14	16	1	0	11	3	(11/14) × 1	(3/14) × 1	0.21	−0.21
15	17	0	1	10	3	(10/13) × 1	(3/13) × 1	−0.77	0.77
16	22	1	1	7	2	(7/9) × 2	(2/9) × 2	−0.56	0.56
17	23	1	1	6	1	(6/7) × 2	(1/7) × 2	−0.71	0.71
Totals		9	21			19.26	10.74	−10.26	+10.26

$$O_i - E_i = \sum_{j=1}^{\#\,failure\ times} \left(m_{ij} - e_{ij}\right)$$

$$O_1 - E_1 = -10.26$$

$$O_2 - E_2 = 10.26$$

$$\text{Log-rank statistic} = \frac{(O_2 - E_2)^2}{Var(O_2 - E_2)}$$

Example: Remission data

EXAMPLE

Expanded Table (Remission Data)

		# failures		# in risk set		# expected		Observed–expected	
j	$t_{(j)}$	m_{1j}	m_{2j}	n_{1j}	n_{2j}	e_{1j}	e_{2j}	$m_{1j}-e_{1j}$	$m_{2j}-e_{2j}$
1	1	0	2	21	21	(21/42) × 2	(21/42) × 2	−1.00	1.00
2	2	0	2	21	19	(21/40) × 2	(19/40) × 2	−1.05	1.05
3	3	0	1	21	17	(21/38) × 1	(17/38) × 1	−0.55	0.55
4	4	0	2	21	16	(21/37) × 2	(16/37) × 2	−1.14	1.14
5	5	0	2	21	14	(21/35) × 2	(14/35) × 2	−1.20	1.20
6	6	3	0	21	12	(21/33) × 3	(12/33) × 3	1.09	−1.09
7	7	1	0	17	12	(17/29) × 1	(12/29) × 1	0.41	−0.41
8	8	0	4	16	12	(16/28) × 4	(12/28) × 4	−2.29	2.29
9	10	1	0	15	8	(15/23) × 1	(8/23) × 1	0.35	−0.35
10	11	0	2	13	8	(13/21) × 2	(8/21) × 2	−1.24	1.24
11	12	0	2	12	6	(12/18) × 2	(6/18) × 2	−1.33	1.33
12	13	1	0	12	4	(12/16) × 1	(4/16) × 1	0.25	−0.25
13	15	0	1	11	4	(11/15) × 1	(4/15) × 1	−0.73	0.73
14	16	1	0	11	3	(11/14) × 1	(3/14) × 1	0.21	−0.21
15	17	0	1	10	3	(10/13) × 1	(3/13) × 1	−0.77	0.77
16	22	1	1	7	2	(7/9) × 2	(2/9) × 2	−0.56	0.56
17	23	1	1	6	1	(6/7) × 2	(1/7) × 2	−0.71	0.71
Totals		9	21			19.26	10.74	−10.26	+10.26

$$O_i - E_i = \sum_{j=1}^{\#\ failure\ times} \left(m_{ij} - e_{ij}\right)$$

$$O_1 - E_1 = -10.26$$

$$O_2 - E_2 = 10.26$$

$$\text{Log-rank statistic} = \frac{(O_2 - E_2)^2}{Var(O_2 - E_2)}$$

Example: Remission data

EXAMPLE

Expanded Table (Remission Data)

		# failures		# in risk set		# expected		Observed–expected	
j	$t_{(j)}$	m_{1j}	m_{2j}	n_{1j}	n_{2j}	e_{1j}	e_{2j}	$m_{1j}-e_{1j}$	$m_{2j}-e_{2j}$
1	1	0	2	21	21	$(21/42)\times 2$	$(21/42)\times 2$	−1.00	1.00
2	2	0	2	21	19	$(21/40)\times 2$	$(19/40)\times 2$	−1.05	1.05
3	3	0	1	21	17	$(21/38)\times 1$	$(17/38)\times 1$	−0.55	0.55
4	4	0	2	21	16	$(21/37)\times 2$	$(16/37)\times 2$	−1.14	1.14
5	5	0	2	21	14	$(21/35)\times 2$	$(14/35)\times 2$	−1.20	1.20

$$O_i - E_i = \sum_{j=1}^{\#\,failure\ times} \left(m_{ij} - e_{ij}\right)$$

$$O_1 - E_1 = -10.26$$

$$O_2 - E_2 = 10.26$$

p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed!

Result

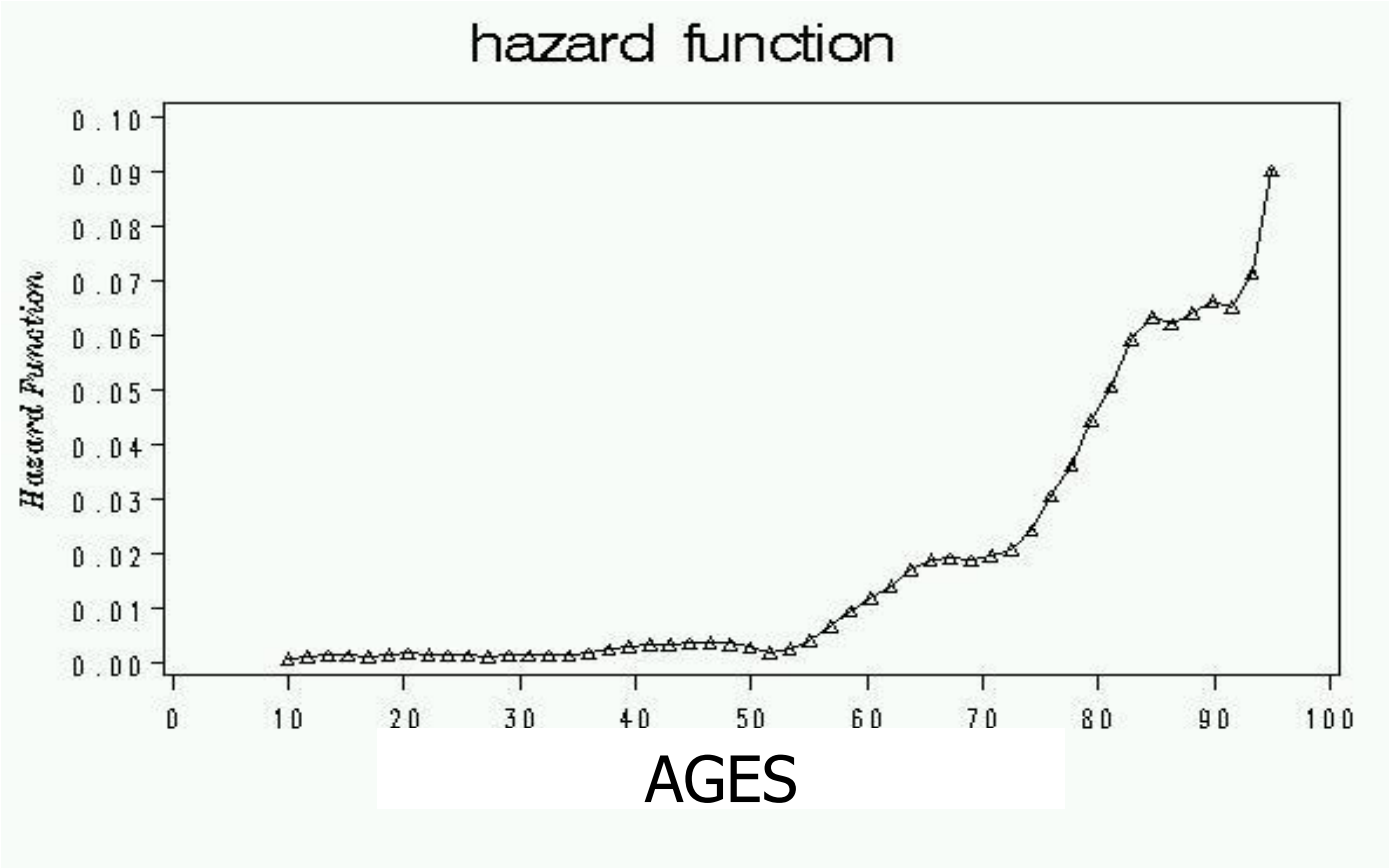
```
> fit
Call:
survdiff(formula = Surv(time, status) ~ treatment)

              N Observed Expected (O-E)^2/E (O-E)^2/V
treatment=1 21        9     19.3      5.46      16.8
treatment=2 21       21     10.7      9.77      16.8

Chisq = 16.8 on 1 degrees of freedom, p = 4.17e-05
```

Hazard Function

Hazard function h(t) (风险函数): the instantaneous rate at time t （在某时点的瞬间死亡率）.

- The hazard function *h(t)* of survival time *T* gives the *conditional failure rate*
- The hazard function is also known as the *instantaneous failure rate, force of mortality,* and *age-specific failure rate*
- *The hazard function gives the risk of failure per unit time during the aging process*

Hazard Function: new concept



Hazard rate is an instantaneous incidence rate

Hazard function

$$h(t) = \lim_{\Delta t \longrightarrow 0} \frac{P(t \le T < t + \Delta t / T \ge t)}{\Delta t}$$

<u>In words:</u> the probability that ***if you survive to t***, you will succumb to the event in the next instant.

$$\text{Hazard from density and survival: } h(\mathrm{t}) = \frac{f(t)}{S(t)}$$

➡ probability density function

➡ Survival function

Derivation (Bayes' rule):

$$h(t)dt = P(t \le T < t + dt / T \ge t) = \frac{P(t \le T < t + dt \,\&\, T \ge t)}{P(T \ge t)} = \frac{P(t \le T < t + dt)}{P(T \ge t)} = \frac{f(t)dt}{S(t)}$$

Hazard function and Survival function

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

$$\text{Hazard from survival: } h(t) = -\frac{d}{dt}\ln S(t)$$

Cox Regression (Cox's Proportional Hazards Model)

- Semi-parametric
- Cox models the effect of predictors and covariates on the hazard rate but leaves the baseline hazard rate unspecified.
- Also called proportional hazards regression
- Does NOT assume knowledge of absolute risk.
- Estimates *relative* rather than *absolute* risk.

The model: Cox regression

Components:

- A baseline hazard function
- A linear function of a set of k fixed covariates that is exponentiated. (=the relative risk)

Risk factor coefficients give hazard ratios (relative risk)

$$\log h_i(t) = \log h_0(t) + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$$

$$h_i(t) = h_0(t) e^{\beta_1 x_{i1} + \ldots + \beta_k x_{ik}}$$

β_l >0表示该协变量是危险因素，越大使生存时间越短

β_l <0表示该协变量是保护因素，越大使生存时间越长

Comparing the survival curves by Age Groups after Adjusting Cellularity using CPHM