EXAMPLE

SNI study: hazard ratio (HR) describes relationship between SNI and T, after controlling for covariates.

Thus, from the example of survival analysis modeling of the social network data, one can obtain a hazard ratio that describes the relationship between SNI and survival time (T), after controlling for the appropriate covariates.

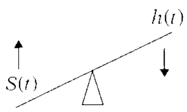
Interpretation of HR (like OR):

HR = 1 $\Rightarrow$ no relationship

HR = 10 $\Rightarrow$ exposed hazard 10 times unexposed

HR = 1/10 $\Rightarrow$ exposed hazard 1/10 times unexposed

The hazard ratio, although a different measure from an odds ratio, nevertheless has a similar interpretation of the strength of the effect. A hazard ratio of 1, like an odds ratio of 1, means that there is no effect; that is, 1 is the null value for the exposure–outcome relationship. A hazard ratio of 10, on the other hand, is interpreted like an odds ratio of 10; that is, the exposed group has ten times the hazard of the unexposed group. Similarly, a hazard ratio of 1/10 implies that the exposed group has one-tenth the hazard of the unexposed group.



$h_E(t) > h_{\bar{E}}(t) \Rightarrow S_E(t) < S_{\bar{E}}(t)$

$$HR = \frac{h_E(t)}{h_{\bar{E}}(t)} > 1 \Rightarrow \frac{S_E(t)}{S_{\bar{E}}(t)} < 1$$

Similarly,

$$HR = \frac{h_E(t)}{h_{\bar{E}}(t)} < 1 \Rightarrow \frac{S_E(t)}{S_{\bar{E}}(t)} > 1$$

Recall that the higher the survival probability at time t, the lower is the corresponding hazard rate, and vice versa.

Therefore, if the hazard rate for an exposed group at time t is higher than that for the unexposed group at the same time, the corresponding survival probability for the exposed group is lower than the corresponding survival probability for the unexposed group. In other words, a hazard ratio greater than 1 corresponds to a ratio of survival probabilities that is less than 1. Similarly, a hazard ratio less than 1 corresponds to a ratio of survival probabilities greater than 1.

EXAMPLE

$$HR = \frac{h_E(t)}{h_{\bar{E}}(t)} = 10 \Rightarrow \frac{S_E(t)}{S_{\bar{E}}(t)} < 1;$$

i.e., $S_E(t) < S_{\bar{E}}(t)$

Thus, for example, if the hazard ratio comparing exposed to unexposed is 10, then the failure rate for the exposed is ten times higher than the failure rate for the unexposed. Consequently, the exposed group must have a lower (though not necessarily 1/10) survival probability than the unexposed. That is, if the hazard ratio is 10, then the exposed group has a poorer survival probability than the unexposed.

Chapters

✓1. Introduction
2. Kaplan–Meier Survival Curves and the Log–Rank Test

This presentation is now complete. We suggest that you review the material covered here by reading the detailed outline that follows. Then do the practice exercises and test.

In Chapter 2 we describe how to estimate and graph survival curves using the Kaplan–Meier (KM) method. We also describe how to test whether two or more survival curves are estimating a common curve. The most popular such test is called the log–rank test.

Detailed Outline

I. What is survival analysis? (pages 4–5)

A. Type of problem addressed: outcome variable is **time until an event occurs.**

B. Assume one event of interest; more than one type of event implies **a competing risk** problem.

C. Terminology: time = survival time; event = failure.

D. Examples of survival analysis:

i. leukemia patients/time in remission

ii. disease-free cohort/time until heart disease

iii. elderly population/time until death

iv. parolees/time until rearrest (recidivism)

v. heart transplants/time until death

II. Censored data (pages 5–8)

A. Definition: don't know exact survival time.

B. Reasons: study ends without subject getting event; lost to follow-up; withdraws.

C. Examples of survival data for different persons; summary table.

III. Terminology and notation (pages 8–14)

A. Notation: T = survival time random variable
t = specific value for T
δ = (0–1) variable for failure/censorship status

B. Terminology: $S(t)$ = survivor function
$h(t)$ = hazard function

C. Properties of survivor function:

- theoretically, graph is smooth curve, decreasing from $S(t) = 1$ at time $t = 0$ to $S(t) = 0$ at $t = \infty$;
- in practice, graph is step function that may not go all the way to zero at end of study if not everyone studied gets the event.

D. Hazard function formula:

$$h(t) = \lim_{\Delta t \to 0} \frac{P\left(t \le T < t + \Delta t \,\middle|\, T \ge t\right)}{\Delta t}$$

E. Hazard function properties:

- $h(t)$ gives instantaneous potential for event to occur given survival up to time t;
- instantaneous potential idea is illustrated by velocity;
- hazard function also called "conditional failure rate";
- $h(t) \ge 0$; has no upper bound; not a probability; depends on time units.

F. Examples of hazard curves:
  i. exponential
  ii. increasing Weibull
  iii. decreasing Weibull
  iv. log normal

G. Uses of hazard function:
  - gives insight about conditional failure rates;
  - identifies specific model form;
  - math model for survival analysis is usually written in terms of hazard function;

H. Relationship of $S(t)$ to $h(t)$: if you know one, you can determine the other.
  - example: $h(t) = \lambda$ if and only if $S(t) = e^{-\lambda t}$
  - general formulae:

$$S(t) = \exp\left[-\int_0^t h(u)du\right]$$

$$h(t) = -\left[\frac{dS(t)/dt}{S(t)}\right]$$

IV. Goals of survival analysis (page 15)

A. Estimate and interpret survivor and/or hazard functions.

B. Compare survivor and/or hazard functions.

C. Assess the relationship of explanatory variables to survival time.

V. Basic data layout for computer (15–19)

A. General layout:

#	t	δ	X_1	$X_2 \cdots X_p$
1	t_1	δ_1	X_{11}	$X_{12} \cdots X_{1p}$
2	t_2	δ_2	X_{21}	$X_{22} \cdots X_{2p}$
•	•	•		•
•	•	•		•
•	•	•		•
j	t_j	δ_j	X_{j1}	$X_{j2} \cdots X_{jp}$
•	•	•		•
•	•	•		•
•	•	•		•
n	t_n	δ_n	X_{n1}	$X_{n2} \cdots X_{np}$

B. Example: Remission time data

VI. Basic data layout for understanding analysis (pages 19–24)

A. General layout:

Ordered failure times ($t_{(j)}$)	# of failures (m_j)	# censored in $(t_{(j)}, t_{(j+1)})$ (q_j)	Risk set $R(t_{(j)})$
$t_{(0)} = 0$	$m_0 = 0$	q_0	$R(t_{(0)})$
$t_{(1)}$	m_1	q_1	$R(t_{(1)})$
$t_{(2)}$	m_2	q_2	$R(t_{(2)})$
•	•	•	•
•	•	•	•
•	•	•	•
$t_{(k)}$	m_k	q_k	$R(t_{(k)})$

Note: k = # of distinct times at which subjects failed; n = # of subjects ($k \le n$); $R(t_{(j)})$, the risk set, is the set of individuals whose survival times are at least $t_{(j)}$ or larger.

B. Example: Remission time data

Group 1 (n = 21, 9 failures, k = 7); **Group 2** (n = 21, 21 failures, k = 12)

C. How to work with censored data:

Use all information up to the time of censorship; don't throw away information.

VII. Descriptive measures of survival experience (pages 24–26)

A. Average survival time (ignoring censorship status):

$$\overline{T} = \frac{\sum_{i=1}^{n} t_i}{n}$$

$\overline{T}$ underestimates the true average survival time, because censored times are included in the formula.

B. Average hazard rate:

$$\overline{h} = \frac{\#\text{ failures}}{\sum_{i=1}^{n} t_i}$$

C. Descriptive measures $\overline{T}$ and $\overline{h}$ give overall comparison; estimated survivor curves give comparison over time.

D. Estimated survivor curves are step function graphs.

E. Median survival time: graphically, proceed horizontally from 0.5 on the Y-axis until reaching graph, then vertically downward until reaching the X-axis.

VIII. Example: Extended remission data (pages 26–29)

A. Extended data adds log WBC to previous remission data.

B. Need to consider **confounding** and **interaction.**

C. Extended data problem: compare survival experience of two groups, after adjusting for confounding and interaction effects of log WBC.

D. Analysis alternatives:

i. stratify on log WBC and compare survival curves for different strata;

ii. use math modeling, e.g., proportional hazards model.

IX. Multivariable example (pages 29–31)

A. The problem: to describe the relationship between social network index (**SNI**) and survival until death controlling for **AGE,** systolic blood pressure (**SBP**), presence or absence of chronic disease (**CHR**), Quetelet's index (**QUET**—a measure of body size), and social class (**SOCL**).

B. Goals:

- to obtain an adjusted measure of effect;
- to obtain adjusted survivor curves for different SNI categories;
- to decide on variables to be adjusted.

C. The data: 13-year follow-up study (1967–1980) of a fixed cohort of n = 170 white males (60+) from Evans County, Georgia.

#	t	δ	SNI	AGE	SBP	CHR	QUET	SOCL
1	t_1	δ_1	SNI_1	AGE_1	SBP_1	CHR_1	$QUET_1$	$SOCL_1$
2	t_2	δ_2	SNI_2	AGE_2	SBP_2	CHR_2	$QUET_2$	$SOCL_2$
•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•
170	t_{170}	δ_{170}	SNI_{170}	AGE_{170}	SBP_{170}	CHR_{170}	$QUET_{170}$	$SOCL_{170}$

X. Math models in survival analysis (pages 32–33)

A. Survival analysis problem is analogous to typical multivariable problem addressed by linear and/or logistic regression modeling: describe relationship of exposure to outcome, after controlling for possible confounding and interaction.

B. Outcome variable (time to event) for survival analysis is different from linear (continuous) or logistic (dichotomous) modeling.

C. Measure of effect in survival analysis: hazard ratio (**HR**).

D. Interpretation of HR: like OR. SNI study: HR describes relationship between SNI and T, after controlling for covariates.

Practice Exercises

True or False (Circle T or F):

T F 1. In a survival analysis, the outcome variable is dichotomous.

T F 2. In a survival analysis, the event is usually described by a (0,1) variable.

T F 3. If the study ends before an individual has gotten the event, then his or her survival time is censored.

T F 4. If, for a given individual, the event occurs **before** the person is lost to follow-up or withdraws from the study, then this person's survival time is censored.

T F 5. $S(t) = P(T > t)$ is called the hazard function.

T F 6. The hazard function is a probability.

T F 7. Theoretically, the graph of a survivor function is a smooth curve that decreases from $S(t) = 1$ at $t = 0$ to $S(t) = 0$ at $t = \infty$.

T F 8. The survivor function at time t gives the instantaneous potential per unit time for a failure to occur, given survival up to time t.

T F 9. The formula for a hazard function involves a conditional probability as one of its components.

T F 10. The hazard function theoretically has no upper bound.

T F 11. Mathematical models for survival analysis are frequently written in terms of a hazard function.

T F 12. One goal of a survival analysis is to compare survivor and/or hazard functions.

T F 13. Ordered failure times are censored data.

T F 14. Censored data are used in the analysis of survival data up to the time interval of censorship.

T F 15. A typical goal of a survival analysis involving several explanatory variables is to obtain an adjusted measure of effect.

16. Given the following survival time data (in weeks),

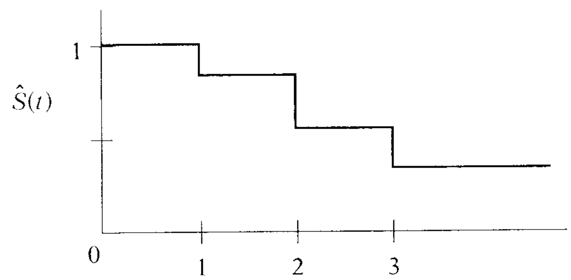
    1, 1, 1+, 1+, 1+, 2, 2, 2, 2+, 2+, 3, 3, 3+, 4+, 5+

    where + denotes censored data, complete the following table:

$t_{(j)}$	$m_{(j)}$	$q_{(j)}$	$R(t_{(j)})$
0	0	0	15 persons survive ≥ 0 weeks
1			
2			
3			

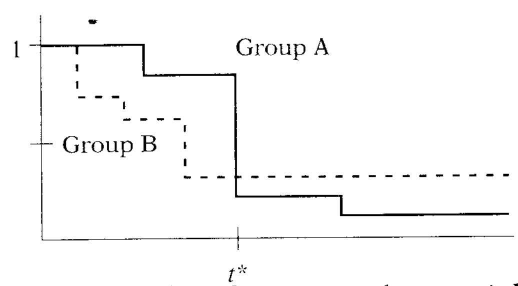
    Also, compute the average survival time ($\bar{T}$) and the average hazard rate ($\bar{h}$) using the raw data (ignoring + signs for $\bar{T}$).

17. Suppose that the estimated survivor curve for the above table is given by the following graph:



    What is the median survival time for this cohort?

Questions 18–20 consider the comparison of the following two survivor curves:



18. Which group has a better survival prognosis **before** time t^*?

19. Which group has a better survival prognosis **after** time t^*?

20. Which group has a longer median survival time?

Test

True or False (Circle T or F):

T F 1. Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable is **time until an event occurs.**

T F 2. In survival analysis, the term "event" is synonymous with "failure."

T F 3. If a given individual is lost to follow-up or withdraws from the study before the end of the study without the event occurring, then the survival time for this individual is said to be "censored."

T F 4. In practice, the survivor function is usually graphed as a smooth curve.

T F 5. The survivor function ranges between 0 and ∞.

T F 6. The concept of instantaneous potential is illustrated by velocity.

T F 7. A hazard rate of one per day is equivalent to seven per week.

T F 8. If you know the form of a hazard function, then you can determine the corresponding survivor curve, and vice versa.

T F 9. One use of a hazard function is to gain insight about conditional failure rates.

T F 10. If the survival curve for group 1 lies completely above the survival curve for group 2, then the median survival time for group 2 is longer than that for group 1.

T F 11. The risk set at six weeks is the set of individuals whose survival times are less than or equal to six weeks.

T F 12. If the risk set at six weeks consists of 22 persons, and four persons fail and three persons are censored by the 7th week, then the risk set at seven weeks consists of 18 persons.

T F 13. The measure of effect used in survival analysis is an odds ratio.

T F 14. If a hazard ratio comparing group 1 relative to group 2 equals 10, then the potential for failure is ten times higher in group 1 than in group 2.

T F 15. The outcome variable used in a survival analysis is different from that used in linear or logistic modeling.

16. State two properties of a hazard function.

17. State three reasons why hazard functions are used.

18. State three goals of a survival analysis.

19. The following data are a sample from the 1967–1980 Evans County study. Survival times (in years) are given for two study groups, each with 25 participants. Group 1 has no history of chronic disease (CHR = 0), and group 2 has a positive history of chronic disease (CHR = 1):

Group 1 (CHR = 0): 12.3+, 5.4, 8.2, 12.2+, 11.7, 10.0, 5.7, 9.8, 2.6, 11.0, 9.2, 12.1+, 6.6, 2.2, 1.8, 10.2, 10.7, 11.1, 5.3, 3.5, 9.2, 2.5, 8.7, 3.8, 3.0

Group 2 (CHR = 1): 5.8, 2.9, 8.4, 8.3, 9.1, 4.2, 4.1, 1.8, 3.1, 11.4, 2.4, 1.4, 5.9, 1.6, 2.8, 4.9, 3.5, 6.5, 9.9, 3.6, 5.2, 8.8, 7.8, 4.7, 3.9

For group 1, complete the following table involving ordered failure times:

	$t_{(j)}$	$m_{(j)}$	$q_{(j)}$	$R(t_{(j)})$	
Group 1:	0.0		0	0	25 persons survived ≥ 0 years
	1.8		1	0	25 persons survived ≥ 1.8 years
	2.2				
	2.5				
	2.6				
	3.0				
	3.5				
	3.8				
	5.3				
	5.4				
	5.7				
	6.6				
	8.2				
	8.7				
	9.2				
	9.8				
	10.0				
	10.2				
	10.7				
	11.0				
	11.1				
	11.7				

20. For the data of Problem 19, the average survival time ($\bar{T}$) and the average hazard rate ($\bar{h}$) for each group are given as follows:

	$\bar{T}$	$\bar{h}$
Group 1:	7.5	.1165
Group 2:	5.3	.1894

a. Based on the above information, which group has a better survival prognosis? Explain briefly.
b. How would a comparison of survivor curves provide additional information to what is provided in the above table?

Answers to Practice Exercises

1. F: the outcome is continuous; time until an event occurs.
2. T
3. T
4. F: the person fails, i.e., is not censored.
5. F: $S(t)$ is the survivor function.
6. F: the hazard is a rate, not a probability.
7. T
8. F: the hazard function gives instantaneous potential.
9. T
10. T
11. T
12. T
13. F: ordered failure times are data for persons who are failures.
14. T
15. T

16.

$t_{(j)}$	$m_{(j)}$	$q_{(j)}$	$R(t_{(j)})$
0	0	0	15 persons survive ≥ 0 weeks
1	2	3	15 persons survive ≥ 1 week
2	3	2	10 persons survive ≥ 2 weeks
3	2	3	5 persons survive ≥ 3 weeks

$$\bar{T} = \frac{33}{15} = 2.2; \quad \bar{h} = \frac{7}{33} = 0.2121$$

17. Median = 3

18. Group A

19. Group B

20. Group A