EXAMPLE (continued)

Group 1 using ordered failure times

$t_{(j)}$	m_j	q_j	$R(t_{(j)})$
$t_{(0)} = 0$	0	0	21 persons survive ≥ 0 wks
$t_{(1)} = 6$	3	$\boxed{1}$	21 persons survive ≥ 6 wks
$t_{(2)} = 7$	1	1	17 persons survive ≥ 7 wks
$t_{(3)} = 10$	1	2	15 persons survive ≥ 10 wks
$t_{(4)} = 13$	1	0	12 persons survive ≥ 13 wks
$t_{(5)} = 16$	1	3	11 persons survive ≥ 16 wks
$t_{(6)} = 22$	1	0	7 persons survive ≥ 22 wks
$t_{(7)} = 23$	1	5	6 persons survive ≥ 23 wks
Totals	9	12	

We now consider, from the table of unordered failure times, person #12 who was censored at 10 weeks, and person #13, who was censored at 11 weeks. Turning to the table of ordered failure times here, we see that these two times are within the third ordered time interval, which starts and includes the 10-week point and ends just before the 13th week. As for the remaining q's, we will let you figure them out for practice.

One last point about the q information. We inserted a row at the top of the data for each group corresponding to time 0. This insertion allows for the possibility that persons may be censored after the start of the study but before the first failure. In other words, it is possible that q_0 may be nonzero. For the two groups of this example, however, no one was censored before the first failure time.

EXAMPLE

Risk Set: $R(t_{(j)})$ is the set of individuals for whom $T \geq t_{(j)}$.

Remission Data: Group 1

$t_{(j)}$	m_j	q_j	$R(t_{(j)})$
$t_{(0)} = 0$	0	0	21 persons survive ≥ 0 wks
$t_{(1)} = 6$	3	1	21 persons survive ≥ 6 wks
$t_{(2)} = 7$	1	1	17 persons survive ≥ 7 wks
$t_{(3)} = 10$	1	2	15 persons survive ≥ 10 wks
$t_{(4)} = 13$	1	0	12 persons survive ≥ 13 wks
$t_{(5)} = 16$	1	3	11 persons survive ≥ 16 wks
$t_{(6)} = 22$	1	0	7 persons survive ≥ 22 wks
$t_{(7)} = 23$	1	5	6 persons survive ≥ 23 wks
Totals	9	12	

The last column in the table gives the **"risk set."** The risk set is not a numerical value or count but rather a collection of individuals. By definition, the risk set $R(t_{(j)})$ is the collection of individuals who have survived at least to time $t_{(j)}$; that is, each person in $R(t_{(j)})$ has a survival time that is $t_{(j)}$ or longer, regardless of whether the person has failed or is censored.

For example, we see that at the start of the study everyone in group 1 survived at least 0 weeks, so the risk set at time 0 consists of the entire group of 21 persons. The risk set at 6 weeks for group 1 also consists of all 21 persons, because all 21 persons survived at least as long as 6 weeks. These 21 persons include the 3 persons who failed at 6 weeks, because they survived and were still at risk just up to this point.

EXAMPLE (continued)

$t_{(j)}$	m_j	q_j	$R(t_{(j)})$
$t_{(0)} = 0$	$\cancel{0}$	$\cancel{0}$	21 persons survive ≥ 0 wks
$t_{(1)} = 6$	$\cancel{3}$	$\cancel{1}$	21 persons survive ≥ 6 wks
$t_{(2)} = 7$	1	1	17 persons survive ≥ 7 wks
$t_{(3)} = 10$	1	2	15 persons survive ≥ 10 wks
$t_{(4)} = 13$	1	0	12 persons survive ≥ 13 wks
$t_{(5)} = 16$	1	3	11 persons survive ≥ 16 wks
$t_{(6)} = 22$	1	0	7 persons survive ≥ 22 wks
$t_{(7)} = 23$	1	5	6 persons survive ≥ 23 wks
Totals	9	12	

$t_{(0)} = 0$	$\cancel{0}$	$\cancel{0}$	21 persons survive ≥ 0 wks
$t_{(1)} = 6$	$\cancel{3}$	$\cancel{1}$	21 persons survive ≥ 6 wks
$t_{(2)} = 7$	$\cancel{1}$	$\cancel{1}$	17 persons survive ≥ 7 wks
$t_{(3)} = 10$	$\cancel{1}$	$\cancel{2}$	15 persons survive ≥ 10 wks
$t_{(4)} = 13$	1	0	12 persons survive ≥ 13 wks
$t_{(5)} = 16$	1	3	11 persons survive ≥ 16 wks
$t_{(6)} = 22$	1	0	7 persons survive ≥ 22 wks
$t_{(7)} = 23$	1	5	6 persons survive ≥ 23 wks
Totals	9	12	

Now let's look at the risk set at 7 weeks. This set consists of seventeen persons in group 1 that survived at least 7 weeks. We omit everyone in the X-ed area. Of the original 21 persons, we therefore have excluded the three persons who failed at 6 weeks and the one person who was censored at 6 weeks. These four persons did not survive at least 7 weeks. Although the censored person may have survived longer than 7 weeks, we must exclude him or her from the risk set at 7 weeks because we have information on this person only up to 6 weeks.

To derive the other risk sets, we must exclude all persons who either failed or were censored before the start of the time interval being considered. For example, to obtain the risk set at 13 weeks for group 1, we must exclude the five persons who failed before, but not including, 13 weeks and the four persons who were censored before, but not including, 13 weeks. Subtracting these nine persons from 21, leaves twelve persons in group 1 still at risk for getting the event at 13 weeks. Thus, the risk set consists of these twelve persons.

How we work with censored data:
Use all informaton up to time of censorship; don't throw away information.

The importance of the table of ordered failure times is that we can work with censored observations in analyzing survival data. Even though censored observations are incomplete, in that we don't know a person's survival time exactly, we can still make use of the information we have on a censored person up to the time we lose track of him or her. Rather than simply throw away the information on a censored person, we use all the information we have.

EXAMPLE

$t_{(j)}$	m_j	q_j	$R(t_{(j)})$
6	3	1	✓ 21 persons
7	1	1	✓ 17 persons
10	1	2	✓ 15 persons
13	1	0	✓ 12 persons
16	1	3	✓ 11 persons
22	1	0	7 persons
23	1	5	6 persons

For example, for the three persons in group 1 who were censored between the 16th and 20th weeks, there are at least 16 weeks of survival information on each that we don't want to lose. These three persons are contained in all risk sets up to the 16th week; that is, they are each at risk for getting the event up to 16 weeks. Any survival probabilities determined before, and including, 16 weeks should make use of data on these three persons as well as data on other persons at risk during the first 16 weeks.

Having introduced the basic terminology and data layouts to this point, we now consider some data analysis issues and some additional applications.

VII. Descriptive Measures of Survival Experience

EXAMPLE

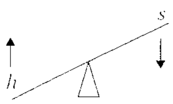
Remission times (in weeks) for two groups of leukemia patients

Group 1 (Treatment) $n = 21$	Group 2 (Placebo) $n = 21$
6, 6, 6, 7, 10,	1, 1, 2, 2, 3,
13, 16, 22, 23,	4, 4, 5, 5,
6+, 9+, 10+, 11+,	8, 8, 8, 8,
17+, 19+, 20+,	11, 11, 12, 12,
25+, 32+, 32+,	15, 17, 22, 23
34+, 35+	
$\bar{T}_1$ (ignoring + 's) 17.1	$\bar{T}_2 = 8.6$
$\bar{h}_1 = \frac{9}{359} = .025$	$\bar{h}_2 = \frac{21}{182} = .115$

$$\text{Average hazard rate } (\bar{h}) = \frac{\#\text{ failures}}{\sum_{i=1}^{n} t_i}$$

We first return to the remission data, again shown in untabulated form. Inspecting the survival times given for each group, we can see that most of the treatment group's times are longer than most of the placebo group's times. If we ignore the plus signs denoting censorship and simply average the survival times for each group we get an average, denoted by T "bar," of **17.1** weeks survival for the treatment group and **8.6** weeks for the placebo group. Because several of the treatment group's times are censored, this means that group 1's true average is even larger than what we have calculated. Thus, it appears from the data (without our doing any mathematical analysis) that, regarding survival, the treatment is more effective than the placebo.

As an alternative to the simple averages that we have computed for each group, another descriptive measure of each group is the **average hazard rate,** denoted as h "bar." This rate is defined by dividing the total number of failures by the sum of the observed survival times. For group 1, h "bar" is 9/359, which equals **.025.** For group 2, h "bar" is 21/182, which equals .115.



As previously described, the hazard rate indicates failure potential rather than survival probability. Thus, the higher the hazard rate, the lower is the group's probability of surviving.

In our example, the average hazard for the treatment group is smaller than the average hazard for the placebo group.

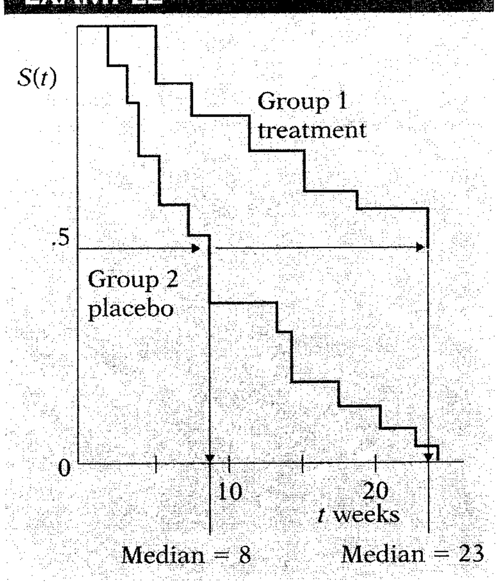
Placebo hazard > treatment hazard: suggests that treatment is more effective than placebo

Thus, using average hazard rates, we again see that the treatment group appears to be doing better overall than the placebo group; that is, the treatment group is less prone to fail than the placebo group.

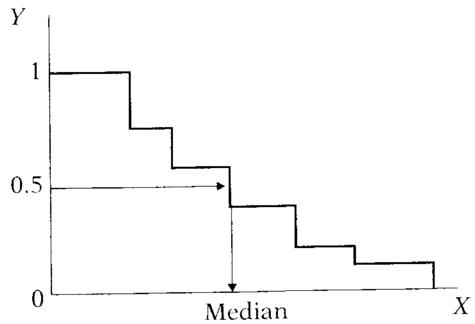
Descriptive measures ($\bar{T}$ and $\bar{h}$) give **overall** comparison; they do not give comparison over time.

The descriptive measures we have used so far—the ordinary average and the hazard rate average—provide overall comparisons of the treatment group with the placebo group. These measures don't compare the two groups at different points in time of follow-up. Such a comparison is provided by a graph of survivor curves.

EXAMPLE



Here we present the **estimated survivor curves** for the treatment and placebo groups. The method used to get these curves is called the Kaplan–Meier method, which is described in Chapter 2. When estimated, these curves are actually step functions that allow us to compare the treatment and placebo groups over time. The graph shows that the survivor function for the treatment group consistently lies above that for the placebo group; this difference indicates that the treatment appears effective at all points of follow-up. Notice, however, that the two functions are somewhat closer together in the first few weeks of follow-up, but thereafter are quite spread apart. This widening gap suggests that the treatment is more effective later during follow-up than it is early on.



Also notice from the graph that one can obtain the median survival time for each group. Graphically, the median is obtained by proceeding horizontally from the 0.5 point on the Y-axis until the survivor curve is reached, as marked by an arrow, and then proceeding vertically downward until the X-axis is crossed at the median survival time.

Median (treatment) = 23 weeks

Median (placebo) = 8 weeks

For the treatment group, the median is 23 weeks, as indicated at the right-hand corner of the top graph. For the placebo group, the median is 8 weeks. Comparison of the two medians reinforces our previous observation that the treatment is more effective overall than the placebo.

VIII. Example: Extended Remission Data

Group 1		Group 2	
t (weeks)	log WBC	t (weeks)	log WBC
6	2.31	1	2.80
6	4.06	1	5.00
6	3.28	2	4.91
7	4.43	2	4.48
10	2.96	3	4.01
13	2.88	4	4.36
16	3.60	4	2.42
22	2.32	5	3.49
23	2.57	5	3.97
6+	3.20	8	3.52
9+	2.80	8	3.05
10+	2.70	8	2.32
11+	2.60	8	3.26
17+	2.16	11	3.49
19+	2.05	11	2.12
20+	2.01	12	1.50
25+	1.78	12	3.06
32+	2.20	15	2.30
32+	2.53	17	2.95
34+	1.47	22	2.73
35+	1.45	23	1.97

Before proceeding to another data set, we consider the remission example data (Freireich et al., *Blood*, 1963) in an **extended form.** The table at the left gives the remission survival times for the two groups with additional information about white blood cell count for each person studied. In particular, each person's log white blood cell count is given next to that person's survival time. The epidemiologic reason for adding log WBC to the data set is that this variable is usually considered an important predictor of survival in leukemia patients; the higher the WBC, the worse the prognosis. Thus, any comparison of the effects of two treatment groups needs to adjust for the possible **confounding effect** of such a variable.