Survival analysis

Patrick Breheny

May 1

Survival analysis

- A very common outcome in medical studies is the time until an event occurs:
  - The time until a patient dies
  - The time until a patient suffers a heart attack
  - The time until a liver transplant patient needs a new liver
  - The time until the recurrence of cancer following treatment
- Data involving such an outcome is often called "time-to-event" data or "failure-time data", and the branch of statistics that deals with analyzing these data is called *survival analysis*

Aplastic anemia study

- The study we are going to look at today involves patients with severe aplastic anemia, a condition in which the bone marrow produces an insufficient number of new blood cells
- These patients were given a bone marrow transplant from a compatible family member
- A common complication of bone marrow transplantation is graft-versus-host disease (GVHD), in which the immune cells produced by the new bone marrow recognize the recipient as a foreign body and mount an attack
- To ward off GVHD, the recipients were randomized to receive one of two drug combinations:
  - Methotrexate (MTX)
  - Methotrexate and cyclosporine (MTX + CSP)

Aplastic anemia study (cont'd)

- The goals of the study were to answer the questions:
  - Which treatment is more effective at preventing GVHD?
  - Which treatment is more effective at saving lives?
- Both outcomes were measured as time-to-event data
- Our goal today is to answer these questions, addressing both statistical significance and clinical impact

What's wrong with a t-test?

- At first, it might seem that the time until an event occurs is continuous, and that we could use methods for continuous data to analyze time-to-event data
- However, there is a fundamental feature of time-to-event data that destroys any attempt to use such methods: the event doesn't always occur!
  - Only 16 out of 46 subjects developed GVHD
  - Only 13 subjects died during the course of the study
- Thus, a lot of the information is *missing*; we know when some of the patients die or develop GVHD, but not all of them

Couldn't we just throw out the missing data?

- It might seem as though you could fix this problem by throwing out the subjects with missing data
- Nothing could be further from the truth!
- Consider an individual from the MTX+CSP group who survived the entire time from transplantation until the end of the study (3.8 years) – this, in a study where many patients died less than a year following transplantation
- Producing such patients is the entire point of the study; throwing this patient out is a very bad idea that goes against the entire principle of the study

A new type of data

- Indeed, there are no good ways to analyze time-to-event data using the methods we have learned already
- It is a fundamentally different type of data than either continuous or categorical data, requiring entirely new approaches:
  - New summary statistics
  - New methods for plotting the data
  - New methods for inference

Censoring and partial information

- Consider the patient who survived for 3.8 years and then the study ended
- It's true that we don't know when this patient would die; however, we know that he survived **at least 3.8 years**
- Any meaningful analysis of time-to-event data has to take this kind of partial information into account; doing so is what survival analysis is all about
- The statistical term for this kind of partial information is that the actual length of time that the patient survived is *censored*

Examples of censoring

There are many kinds of things that can lead to censoring:

- The end of the study
- Patient moves and the investigators lose contact with them
- Death: for example, when studying graft-versus-host disease, the patient may die before we see whether or not they develop GVHD

The data

- Time-to-event outcomes consist of two pieces of information:
  - The length of time that the patient was on the study
  - Whether or not the time period ended with the event in question (as opposed to some sort of censoring event)
- For example, in the anemia data:

Therapy	Time	GVHD
MTX+CSP	3	No
MTX+CSP	10	Yes
MTX+CSP	12	No
MTX+CSP	16	Yes
MTX+CSP	22	Yes
	...	

Time on study

- It is worth mentioning what exactly is being measured by the "time" variable
- Rarely is it actual "calendar time"
- Instead, the "clock" starts "ticking" for each patient individually at the moment when they enter the study (for the anemia data, when they receive the bone marrow transplant)
- In reality, these transplants all occur on different dates
- However, by starting the measurement of time when transplantation occurs, we create an abstract setting in which all patients receive transplants at the same time (time=0)
- Time measured in this way is sometimes called "time on study"

Functions of time

- The key concept in thinking about survival data is to think of everything that occurs as a function of time (time on study)
- Random variables that vary with time are often called *processes*
- As time progresses forward, several fundamental things change:
  - The events that we are studying occur
  - Censoring occurs
  - The amount of people in our study changes

Notation

- Analyzing time-to-event data requires keeping track of two processes:
  - $n(t)$, the number of subjects in the study at time t; more precisely, the number *at risk* for the event at time t
  - $d(t)$, the number of events that occur at time t
- These processes occur separately in the two groups, and subscripts are often used to distinguish the processes $\{n_1(t), n_2(t), d_1(t), d_2(t)\}$

Notation: example

For example, in the anemia data:

Therapy	Time	GVHD
MTX+CSP	3	No
MTX+CSP	10	Yes
MTX+CSP	12	No
MTX+CSP	16	Yes
MTX+CSP	22	Yes
	...	

t	$n(t)$	$d(t)$
0	22	0
3	22	0
4	21	0
10	21	1
11	20	0
12	20	0
16	19	1
22	18	1
	...	

Notation: example (cont'd)

In the MTX only group,

Therapy	Time	GVHD
MTX	9	Yes
MTX	11	Yes
MTX	12	Yes
MTX	20	Yes
MTX	20	Yes
MTX	25	Yes
MTX	25	Yes
MTX	25	No
MTX	28	Yes
MTX	28	Yes
	...	

t	$n(t)$	$d(t)$
0	24	0
9	24	1
11	23	1
12	22	1
20	21	2
25	19	2
28	16	2
	...	

The survival function

- Summary statistics for time-to-event data revolve around the *survival function*
- Like so many other things in survival analysis, the survival function is a function of time, and is defined as the probability of the event in question not occurring by time t (*i.e.*, the patient surviving until time t or later)
- The survival function is usually denoted by $S(t)$
- So, in the GVHD example, if we say that $S(10) = .95$, this means that there is only a 5% chance of developing GVHD by day 10 (or equivalently, a 95% chance of surviving GVHD-free until day 10)

Estimating $S(t)$: the challenge

- It might seem easy to calculate the observed $S(t)$ in our sample: if 20% of our subjects have developed GVHD by day 30, then $\hat{S}(30) = .8$
- Recall, however, the fundamental challenge of time-to-event data: we don't get to observe the time at which everyone develops GVHD
- If, say, a patient is censored at day 3, we have no idea whether they would have developed GVHD by day 30 or not

Estimating $S(t)$: the solution

- The solution to this problem was proposed by Kaplan and Meier in 1958
- Their insight was to recognize that at time t, the observed probability of the event occurring is

$$\frac{d(t)}{n(t)}$$

or conversely, the probability of the event not occurring is

$$\frac{n(t) - d(t)}{n(t)}$$

Estimating $S(t)$: the solution (cont'd)

- Thus, we can estimate $S(t)$ by starting at $t = 0$, where $S(0) = 1$, and moving forward in time, decreasing $S(t)$ by $d(t)/n(t)$ as we go
- Expressing this as a formula,

$$\hat{S}(t) = \prod_i \frac{n(t_i) - d(t_i)}{n(t_i)}$$

  where the Π is analogous to the Σ notation for sums, but refers to products, and the t_i's are all of the times at which the event occurred prior to t
- The method of estimating $S(t)$ is called the *Kaplan-Meier estimator*

Example

For example, in the MTX+CSP group,

t	$n(t)$	$d(t)$
0	22	0
10	21	1
16	19	1
22	18	1
	...	

t	$\frac{n(t)-d(t)}{n(t)}$	$\hat{S}(t)$
0	1	1
10	$20/21$	.952
16	$18/19$	.902
22	$17/18$	.852
	...	

Example (cont'd)

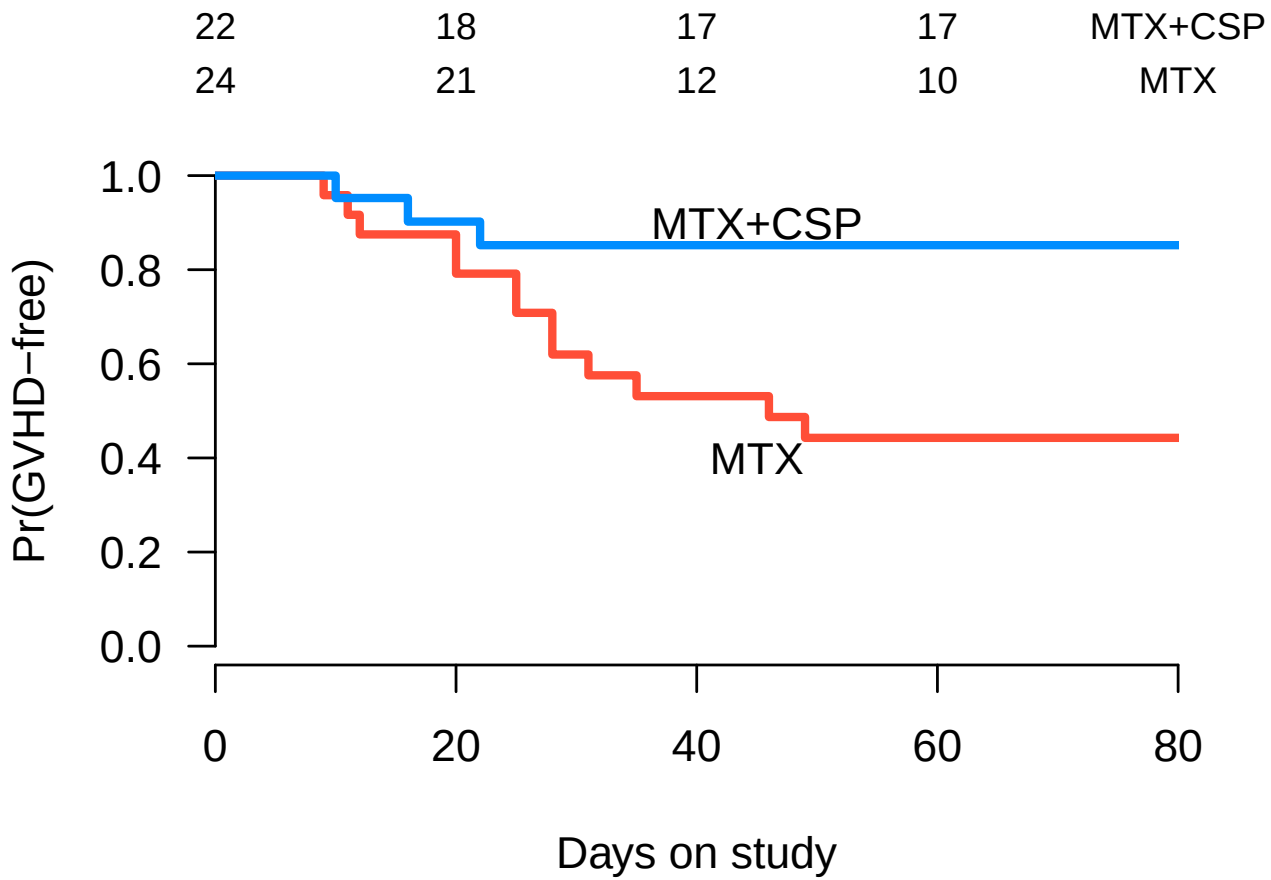
In the MTX group,

t	$n(t)$	$d(t)$
0	24	0
9	24	1
11	23	1
12	22	1
20	21	2
25	19	2
28	16	2
	...	

t	$\frac{n(t)-d(t)}{n(t)}$	$\hat{S}(t)$
0	$24/24$	1
9	$23/24$	.958
11	$22/23$	.917
12	$21/22$	.875
20	$19/21$	.792
25	$17/19$	.708
28	$14/16$	.620
	...	

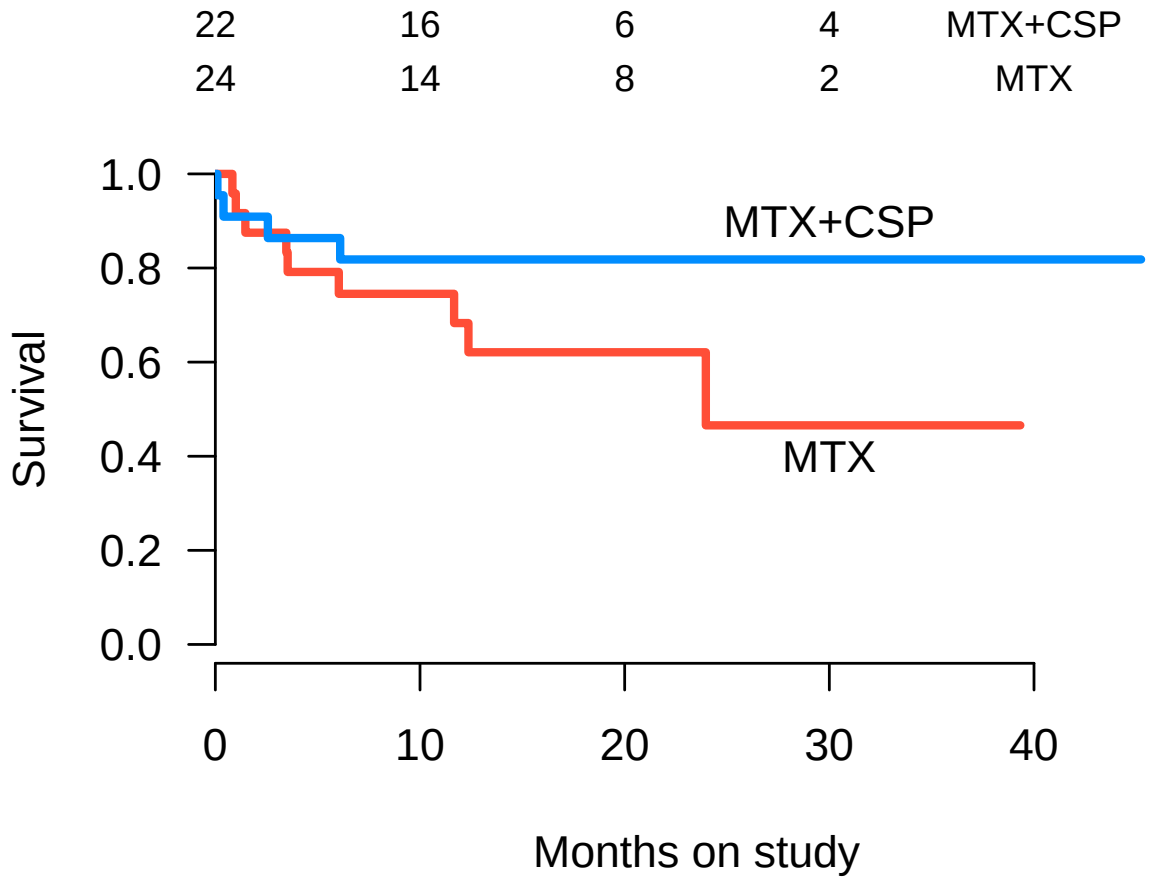
Kaplan-Meier curve: GVHD

The result of all these calculations is usually summarized in a plot called a *Kaplan-Meier curve*:



Kaplan-Meier curve: Survival

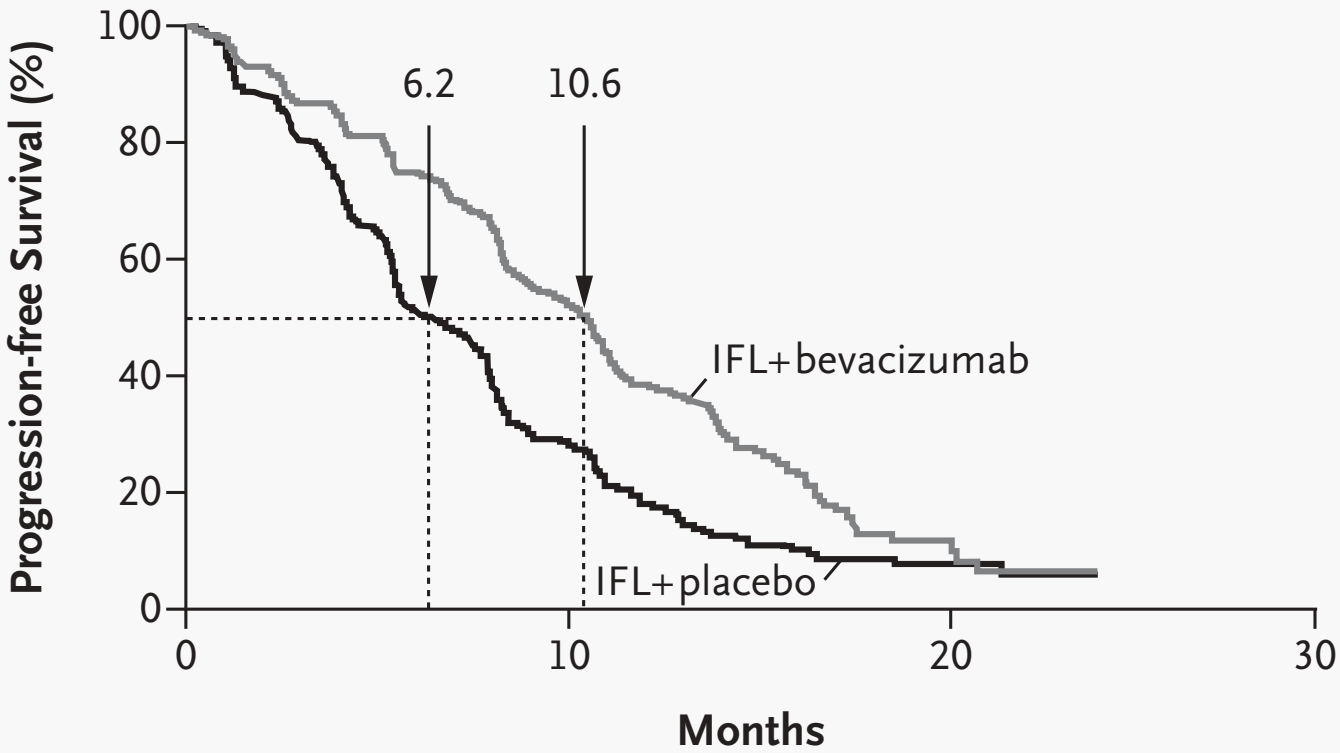
We could go through the same process again, only treating deaths as the event of interest:



Summary statistics

- Summary statistics for time-to-event data are derived from the Kaplan-Meier estimate
- For example, the authors of this study reported "2 year survival rates of 82% and 60%" in the two groups
- In addition, a *median survival time* is also widely reported
- For the MTX group, the median survival time is 24 months; however, a median survival time for the MTX+CSP group cannot be calculated because the Kaplan-Meier curve never drops below .5

Kaplan-Meier curve: Avastin study

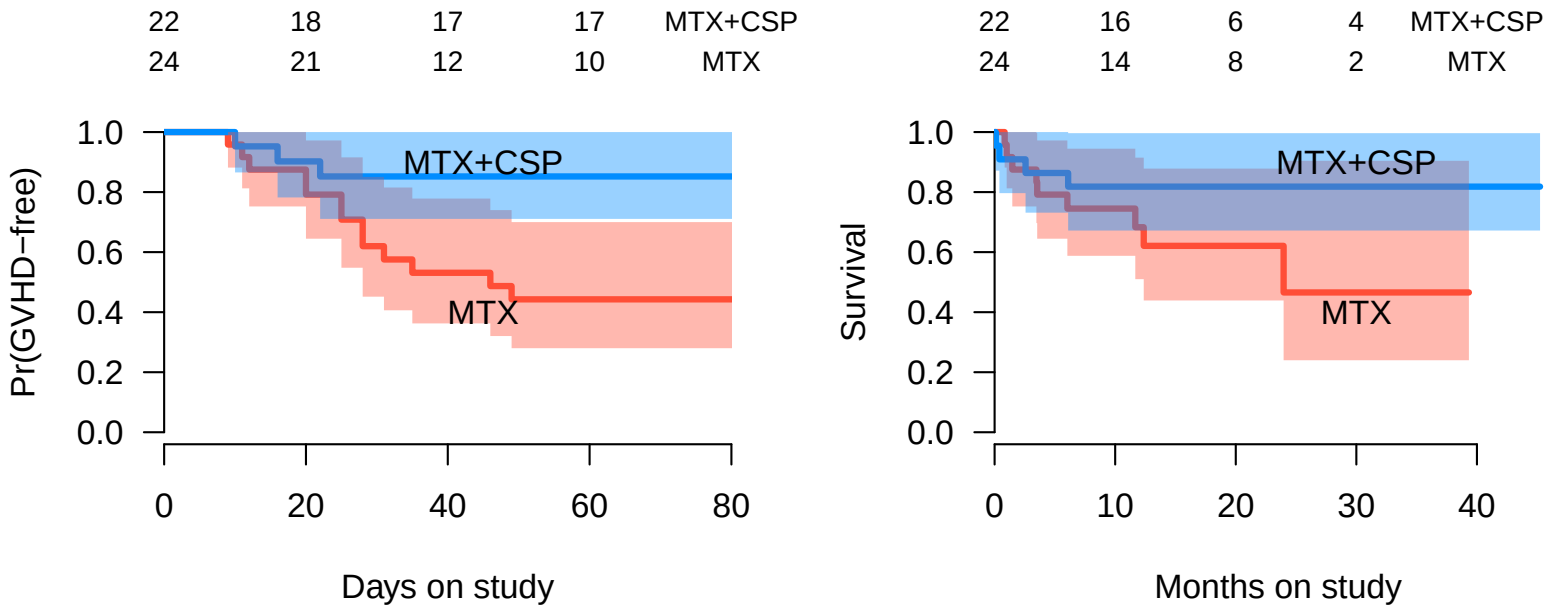


No. at Risk

IFL+bevacizumab	402	269	143	36	6	0
IFL+placebo	411	225	73	17	8	0

Confidence intervals

In addition, we can calculate confidence intervals for the curves:



The log-rank test

- Having calculated these survival curves, a natural question is: are they significantly different from each other?
- The standard hypothesis test for answering this question is called the *log-rank test* and was first proposed in 1966 by Nathan Mantel
- This approach relies on the same idea as the Kaplan-Meier estimator of starting at time 0 and progressing forward through time, focusing on the times at which events occur

Constructing a 2x2 table

- For each time at which the event in question occurs, we can construct a 2x2 table
- For example, a patient in the MTX group contracted GVHD on day 9:

	GVHD	
	Yes	No
MTX	1	23
MTX+CSP	0	21

- We could perform a χ^2-test on these data, but with only one observed case of GVHD, it would have very low power

Many 2x2 tables

- But of course, we don't just have one observed case of GVHD
- The very next day (day 10), we have:

	GVHD	
	Yes	No
MTX	0	23
MTX+CSP	1	20

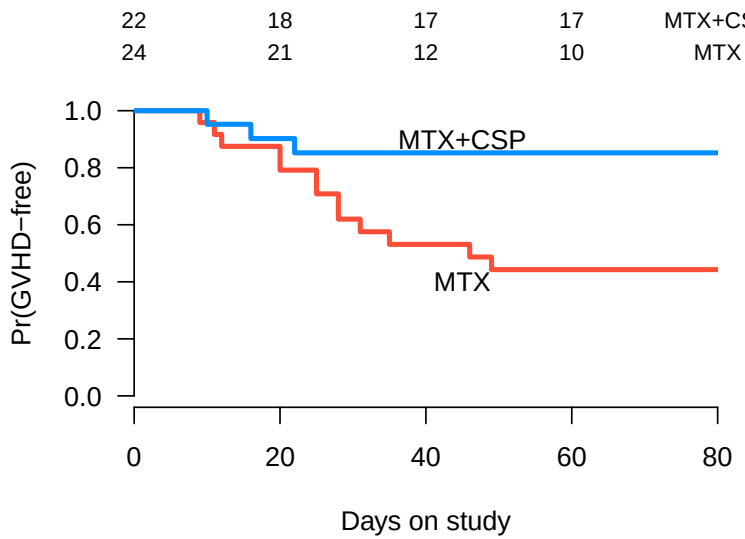
- Note that the evidence at day 10 favors MTX, while the evidence at day 9 favored MTX+CSP

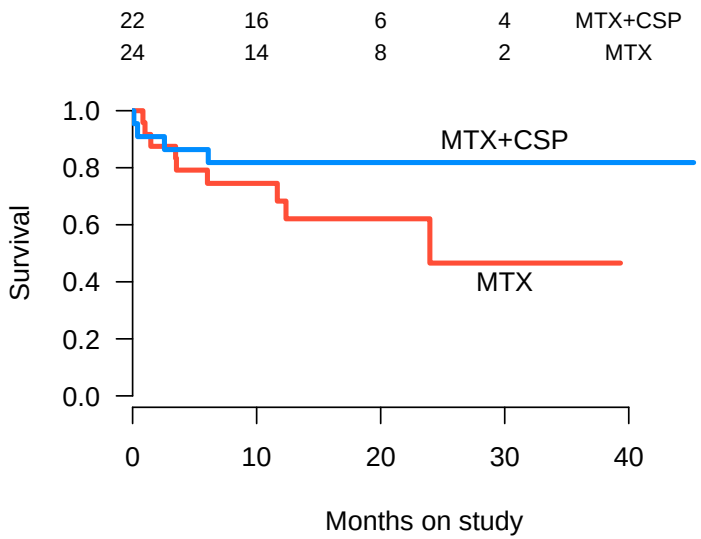
Combining 2x2 tables

- The beauty of the log-rank test lies in how it combines all of these 2x2 contingency tables
- In a sense, this is like calculating the weighted average of all the test statistics from the individual tests
- However, the exact details get somewhat complicated, and we will not go over them in this course
- We'll just look at some log-rank test results and use R (in lab) to carry out the test

Kaplan-Meier curve: GVHD

For GVHD, the p-value of the log-rank test is .01; for survival, the p-value is 0.16:





Kaplan-Meier curves and p-values

- This last example emphasizes an important point with respect to the interpretation of Kaplan-Meier curves
- Kaplan-Meier curves often have long flat regions
- These regions catch the eye, but remember that they represent the absence of data: no events are occurring in these regions
- Accordingly, these regions receive no weight from the log-rank test
- Final comment: the amount of variability present in the Kaplan-Meier estimate of a survival curve is not intuitive, so it is important to look at confidence intervals and p-values in addition to the curves themselves

Summary

- Time-to-event analysis is complicated by the presence of censoring, which produces partially missing outcomes
- The Kaplan-Meier curve is an elegant way to estimate survival using that partial information appropriately; know how to calculate $\hat{S}(t)$
- Two-year/five-year/etc. survival probabilities and median survival times are also common survival summaries, and can be determined from a Kaplan-Meier curve
- The log-rank test can be used to test the null hypothesis of equal survival curves