**Survival Basics**

**Concept Activity**

The goal of this activity is to understand how to properly account for censoring when estimating survival and what a Kaplan-Meier curve tells you.

Let’s use the same imaginary tiny cancer treatment study example that Motulsky uses in Chapter 5 of our textbook.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Participant ID** | **Starting Date** | **Ending Date** | **Time to Event**  **(years)** | **What Happened** |
| A | Feb. 7, 1998 | Mar. 2, 2002 | 4.07 | Died of cancer |
| B | May 19, 1998 | Nov. 30, 2004 | 6.54 | Moved away (lost to follow-up) |
| C | Nov. 14, 1998 | Apr. 3, 2000 | 1.39 | Died of cancer |
| D | Mar. 4, 1999 | May 4, 2005 | 6.17 | Alive at end of study |
| E | June 15, 1999 | May 4, 2005 | 5.89 | Died of cancer |
| F | Dec. 1, 1999 | Sept. 4, 2004 | 4.76 | Died of cancer |
| G | Dec. 15, 1999 | Aug. 15, 2003 | 3.67 | Died in car accident |

1. How many participants were enrolled in the study?
2. How many participants had been lost to follow-up at the end of the study?
3. What defines the starting date for a participant in this study?

Survival studies are often called **time-to-event studies**.

1. What is the event of interest in this study?
2. How many of the enrolled participants are known to have had the event?
3. How many participants were censored? Describe why they were censored.

**Deleting Censored Cases**

Think about the participants with censored observations, and specifically about participants B and D.

1. What information would you lose about the probability of survival for these participants if you simply left participants B and/or D out of the analysis? In what direction do you *expect* that this would bias the result?

Let’s explore what *actually* happens when study participants with censored data are deleted from the data set. The table below includes only the study participants who died of cancer, in order by time.

|  |  |  |
| --- | --- | --- |
| **Participant ID** | **Time to Event (years)** | **What Happened** |
| C | 1.39 | Died of cancer |
| A | 4.07 | Died of cancer |
| F | 4.76 | Died of cancer |
| E | 5.89 | Died of cancer |

1. Use the data in the table above to fill in the survival table below. The first few lines have been done for you. The descriptions of the column titles are described below the table.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Time (years)** | **Number at Risk** | **Number Who Had the Event** | **Percent Who Had the Event** | **Percent Who Did NOT Have the Event** | **Cumulative Survival (%)** |
| 0.00 | 4 | 0 | 0% | 100% | 100% |
| 1.39 | 4 | 1 | 25% | 75% | 75% |
| 4.07 | 3 | 1 | 33.3% | 66.7% | 50% |
| 4.76 |  |  |  |  |  |
| 5.89 |  |  |  |  |  |

Terminology:

* **Number at risk**. This is the number of enrolled participants who are still being followed and are known to be alive (and therefore at risk of the event of interest) immediately before the designated time.
* **Number who had the event**. This is the number of participants who had the event of interest at the designated time.
* **Percent who had the event**. This is the *number of participants who had the event* divided by the *number at risk*.
* **Percent who did not have the event**. This is 100% minus the *percent who had the event*.
* **Cumulative survival (%)**. This is the percent who survived the first time, multiplied by the percent who survived the second time, multiplied by the percent who survived the third time, etc.
  + For example, if half survive the first time, and half of those left survive the second time, then the overall survival after the second time is 25%.

1. Sketch a plot of the cumulative percent survival vs. time. [*Online students*: Please sketch the plot on paper in order to understand how it works, but you do not have to complete this question on the computer unless you wish to. Possible completely-optional approaches include using Excel, or scanning your hand-drawn plot to a PDF and pasting it in.]
2. What do the drops in the curve represent? How many drops are there?
3. What is the estimated median survival time (i.e., the time when the survival curve reaches 50%)?
4. What is the estimated five-year [cumulative] survival?
5. What is the estimated probability of surviving beyond 5.89 years?

**Including Censored Cases**

Now, let’s explore what happens when patients whose data have been censored are kept within the data set and *properly* accounted for, using the Kaplan-Meier method, the accepted method for analyzing survival data. The table below includes all study participants, in order by time.

|  |  |  |
| --- | --- | --- |
| **Participant ID** | **Time to Event**  **(years)** | **What Happened** |
| C | 1.39 | Died of cancer |
| G | 3.67 | Died in car accident |
| A | 4.07 | Died of cancer |
| F | 4.76 | Died of cancer |
| E | 5.89 | Died of cancer |
| D | 6.17 | Alive at end of study |
| B | 6.54 | Moved away (lost to follow-up) |

1. Use the data in the table above to fill in the survival table below. The first few lines have again been done for you.

*Note.* Be careful about the censored participants – they don’t have the event, but they are no longer in the risk pool after they have been censored.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Time (years)** | **Number at Risk** | **Number Who Had the Event** | **Percent Who Had the Event** | **Percent Who Did NOT Have the Event** | **Cumulative Survival (%)** |
| 0.00 | 7 | 0 | 0% | 100% | 100% |
| 1.39 | 7 | 1 | 14.3% | 85.7% | 85.7% |
| 3.67 | 6 | 0 | 0% | 100% | 85.7% |
| 4.07 | 5 | 1 | 20% | 80% | 68.6% |
| 4.76 |  |  |  |  |  |
| 5.89 |  |  |  |  |  |
| 6.17 |  |  |  |  |  |
| 6.54 |  |  |  |  |  |

1. Sketch a plot of cumulative survival percent vs. time (i.e., a Kaplan-Meier curve). Mark the censored patients as large black dots on the curve. [*Online students*: Please sketch the plot on paper in order to understand how it works, but you do not have to complete this question on the computer unless you wish to. Possible completely-optional approaches include using Excel, or scanning your hand-drawn plot to a PDF and pasting it in.]
2. What do the drops in the curve represent? How many drops are there?
3. What is the estimated median survival time (i.e., the time when the survival curve reaches 50%)?
4. What is the estimated five-year [cumulative] survival?
5. What is the estimated probability of surviving beyond 5.89 years?
6. Why doesn’t the survival drop to 0% at the end of the plot?

**Excluding vs. Including**

1. Compare the median survival time estimates from the two approaches above. What is the effect of excluding censored observations?
2. Compare the five-year survival estimates from the two approaches above. What is the effect of excluding censored observations?
3. Compare the estimated probability of surviving longer than 5.89 years obtained from the two approaches above. What is the effect of excluding censored observations?
4. Suppose that your colleague wants to found an article that has a survival curve in it. They aren’t sure on how to interpret the curve. How would you explain to your colleague what information a survival curve provides?

**EXTENSIONS**

Survival analysis methods aren’t restricted to cases where the event is death.

1. Think of an example in your field of a time-to-event situation. Be sure to specify what the event is.