

*BIOS 522: Survival Analysis Methods*

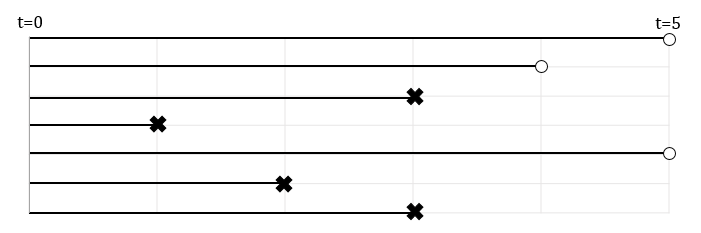
**Activity 1:**

**Introduction to time-to-event data**

*This week, we provided an overview of methods for time-to-event data. We studied examples of failure time events, time origins, and time scales. We defined and characterized right censoring. We reviewed the analysis of continuous, binary, and person-time data.*

Problem 1. Sample cohort data

Below is hypothetical survival data for 7 cohort study participants. Each participant is enrolled at time and followed for up to *years*. An “X” indicates the time when a participant was observed to die. An “O” indicates the time when a participant was censored due to loss to follow-up or end of study. You can assume that individuals who were censored did not have the event at the censoring time (e.g. the first participant survived beyond ).



Use this hypothetical data to calculate the following quantities. Answers can be reported as raw fractions or using a calculator. The answers to some questions may be “not calculable.” Be prepared to share your answer with the group.

1. Write out the data in notation.
2. Write out the data in shorthand notation (e.g., 1, 2+, 3, 4+)

(1,2, 3, 3,4+,5+,5+)

1. Calculate the probability of death at 2 years after enrolment.

Answer: 2/7

1. Calculate the probability of death at 5 years after enrolment.

Answer: not calculable. Though it lies in the range 4/7 to 5/7

1. Calculate the incidence rate during the study period.

Answer: Person time = 5+4+3+1+5+2+3 = 23

Events = 4

Incidence rate = 4/23 or 0.17 events per PY

1. Calculate the mean survival time.

Answer: not calculable.

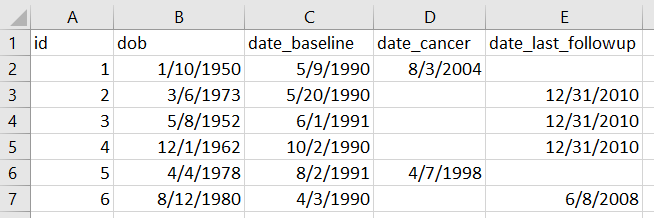
1. Calculate the median survival time:

Answer: 3 years.

Problem 2. More sample cohort data

Below is an excerpt from a spreadsheet of data for 6 hypothetical cohort study participants. Cancer-free participants are invited to enroll in the prospective study during 1990-1991 and are followed over a period of 20 years. The outcome of interest is diagnosis of cancer.

* “id” is the unique participant identifier
* “dob” is the date of birth (all dates in mm/dd/yyyy format)
* “date\_baseline” is the date of enrolment into the study
* “date\_cancer” is the date of cancer diagnosis for participants diagnosed with cancer during the follow-up period
* “date\_last\_followup” is the end of the study (12/31/2010) or date of loss to follow-up (e.g. participant id=6, who moves out of the study region before the end of the study).



1. Describe how you would create a right-censored follow-up time variable for time since enrolment into the study. (*It is not necessary to calculate the follow-up times exactly. Just describe the process*.)

Calculate difference between date cancer and date of baseline. For these individuals, set the event indicator to 1. For all other individuals (censored at last date of follow-up), calculate the difference between date of baseline and date of last follow-up, and set the event indicator to 0.

1. What is a limitation of using time since enrolment in the analysis? Think about the underlying scientific question.

Time since enrolment is not a biologically meaningful quantity that predicts cancer risk. Age, on the other hand, is an important predictor. Consider calculating a variable for age at enrolment as a covariate in models.

Problem 3. Incidence rates

What is the difference between an incidence rate and cumulative incidence? Describe their units. Can each exceed one?

A cumulative incidence is a probability. It is bounded between 0 and 1. It does not have units of time, but it is calculated assuming a particular duration of follow-up (e.g. cumulative incidence of cancer progression 5 years after diagnosis).

Incidence rates are calculated based on person-time, as they are calculated from the number of events divided by person-time follow up. They will have units like per 100,000 person-years. They are rates, so they are necessarily positive, but are not bounded above by 1.