What is Kaplan Meier curve?

* Graph plot of the survival probability against time-to-events, showing the probability of a subject will survive up to time
* Non-parametric estimate which does not make assumption about the distribution of data
* Estimate from data that could be censored, do not experience the events

Censored data (the data is incomplete without end time for events)

* Not experience the event before end of Kaplan Meier analysis
* Follow up is missing or withdraw from the analysis

If do not include the censored data, the sample size of the study may become small

For Kaplan Meier curve

In specific question:

What is the difference in survival of **Malignant neoplasm of breast (disorder)** treated with **Paclitaxel 100 MG Injection** compared to **100 ML Epirubicin Hydrochloride 2 MG/ML Injection**?

Here the event in interest is the death of the patient, we assume the death is caused by the cancer.

1. Get the time range for the Kaplan Meier analysis. (Here let start time be the earliest time for a patient got treatment and end time as the time the data extracted)
2. Clinical outcome (alive or death) is used to get status, if death then status as 1 otherwise status as 0
3. Get start time the patient got treatment here is the earliest time the patient got specified medication(data must have start time)
4. Get end time from patient table (DEATHDATE), if alive then the end time is the end time of the Kaplan Meier analysis,
5. Then compute duration and get the data with columns: duration, status, other group variables to do plot.

|  |  |  |
| --- | --- | --- |
| Duration/Time | Status/events | group |
|  |  |  |

1. And then for each time point for a specific group, we count the number of event censored whose status value is 0; and the number of event occurred which are events with status 1(patient died); The number of events at risk is equal to the number of events at risk at last time point minus the number of event censored minus and the number of event occurred

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Time(t) | Number of event censored() | Number of event occurred() | Number of event at risk() | Survival probability( ) |
|  |  |  |  |  |

According to formula to calculate the survival probability

= where, = - -

|  |  |  |
| --- | --- | --- |
| Time(t) | Survival probability() | group |
|  |  |  |

Reference:

<https://yuzar-blog.netlify.app/posts/2021-01-03-survival-analysis-1-a-gentle-introduction-into-kaplan-meier-curves/>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3059453/>

<https://towardsdatascience.com/kaplan-meier-curve-explained-9c78a681faca>