Novel Approaches for visualising Time to Event Health Data

Victor Pacifique Rwandarwacu, Ranganath Shyamsundar[1]

[1] School of Mathematical and Statistical Sciences, University of Galway

Background and objectives

Time-to-event or survival Analysis is the analysis of data in the form of times from a well-defined time origin until the occurrence of some particular event or end point1[1]. Survival data are generally asymmetric and censored, which requires the use of specific approaches for analysis and visualisations, such as this survival function, Kaplan Meier(KM) estimator and plot.

The survival function $S(t)$ is the probability that the survival time is greater than or equal to time (t) which is the observed value of random variable T with distribution function $F(t)$[2].

$$S(t) = \mathrm{P}(T \geqslant t) = 1 - F(t)$$

$$F(t) = \mathrm{P}(T < t) = \int_0^t f(u)\mathrm{d}u$$

The Kaplan Meier estimate of the survival function at kth interval is given by:

$$\hat{S}(t) = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j} \right)$$

For $t_{(k)} \leqslant t < t_{(k+1)}, k = 1, 2, \ldots, r$, with $\hat{S}(t) = 1$ for $t < t_{(1)}$,where t_{r+1} is taken to be ∞ d_j denotes the number of deaths in this interval, n_j is the number of individuals alive just before $t_{(j)}$ and d_j deaths at $t_{(j)}$.

Survival Ratio, a robust approach for comparing survival distributions [3], is defined by:

$$R(t) = \frac{S_1(t)}{S_2(t)}$$

This project explores the use of novel informative visualisations of time-to-event data, specifically comparing survival curves of different covariates or treatments in a trial.

Data Sources and Datasets

The dataset is from the NIH National Cancer Institute , TCGA Program on a project called “Breast invasive carcinoma (BRCA)”. It contains information about: demography, exposure , Family History(regarding cancer), Follow up, Molecular Test, other Clinical Attribute, pathology detail,and Treatment of female Breast cancer patients diagnosed and followed up for different outcomes.For demonstration, our analysis focuses on Survival outcomes by pathologic stages [4]

Proposed Approach

Table 1 demonstrates the section of survival function and the change of the number of people at risk on each time interval.

Table 1: **BRCA data survival**

Time_Yrs	Survival_Prob	n.risk	Std.Error	Lower.95CI	Upper.95CI
5.256673	0.8131862	233	0.0182546	0.7781836	0.8497633
5.275838	0.8096506	230	0.0185144	0.7741642	0.8467637
5.456537	0.8060035	222	0.0187868	0.7700105	0.8436791
5.500342	0.8023399	220	0.0190553	0.7658481	0.8405705
5.741273	0.7985009	209	0.0193470	0.7614679	0.8373351
5.823409	0.7946058	205	0.0196408	0.7570282	0.8340488

Figure 1, 2 and 3 highlight different approaches of visualising the estimated survival function.

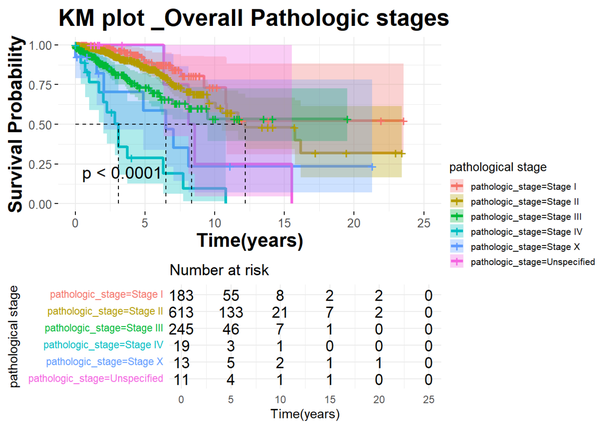

Figure 1: *KM plot_overall pathologic stages*

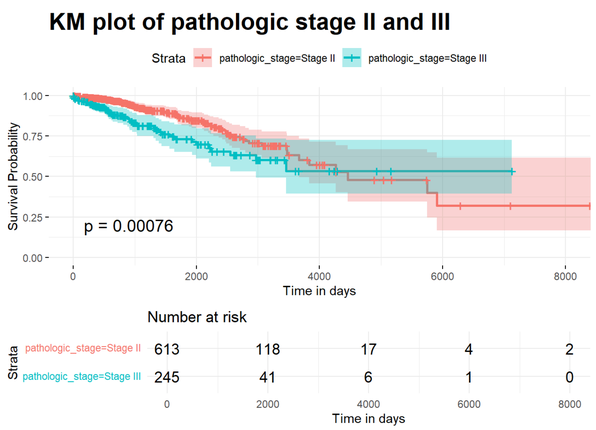

Figure 2: *KM plot Of Pathologic stage II and III*

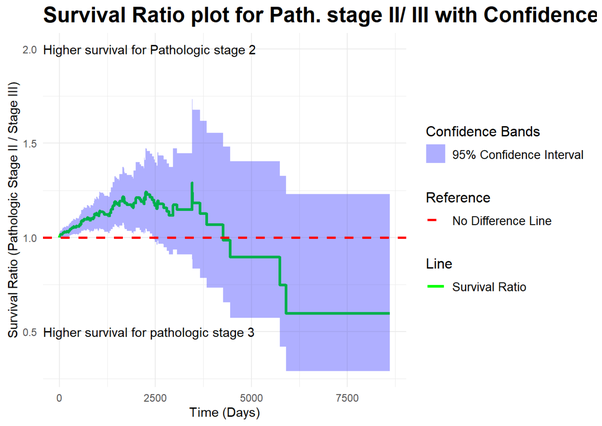

Figure 3: *Survival Ratio plot for Path. stage II/ III with 95% C.I*

Next Steps

-Visualize survival differences between independent groups, incorporating confidence intervals to assess variability and significance.

-Generate survival ratio plots for paired data, using permutation envelopes as reference bands to evaluate deviations and provide robust comparisons.

-Compare survival distributions across more than two groups utilizing non-parametric statistical methods to identify significant differences.

GitHub

The code and dataset for this project can be at GitHub repository through this link : https://github.com/rwandarwacu1/Msc_thesis_survival

References

1. David Collett, Modelling survival data in medical research , Fourth Ed.↩

2. Peace, Karl E.. Design and Analysis of Clinical Trials with Time-to-Event Endpoints (Chapman & Hall/CRC Biostatistics Series) (p. 74). CRC Press. Kindle Edition.↩

3. J.Newell et.al, Survival ratio plots with permutation envelopes in survival data problems, https://doi.org/10.1016/j.compbiomed.2005.03.005↩

4. <TCGA-BRCA , https://portal.gdc.cancer.gov/projects/TCGA-BRCA>↩