Report on Survival Analysis of Breast Cancer Patients

-Sujal Yadav

Introduction

Breast cancer is one of the most common cancers among women worldwide, and understanding the factors that influence survival rates is crucial for improving patient outcomes. Survival analysis is a statistical method that allows us to estimate the time until an event occurs, such as recurrence or death, and to identify factors that may influence this time. In this report, we perform a comprehensive survival analysis on a dataset of breast cancer patients, focusing on recurrence-free survival time. The analysis includes Kaplan-Meier estimation and model fitting, considering various factors such as tumour grade, hormonal therapy, and menopausal status.

The dataset contains information on several variables, including age, menopausal status, tumour size, tumour grade, number of positive lymph nodes, progesterone receptor status, oestrogen receptor status, and hormonal therapy status. The outcome variable is the recurrence-free survival time, with a censoring indicator showing whether an event (recurrence) has occurred or not.

Data Overview

The dataset contains the following key variables:

- **Recurrence-Free Survival Time:** Time in days until recurrence or censoring.
- **Status:** Indicator of whether an event (recurrence) has occurred (1) or if the data is censored (0).
- **Age:** Age of the patient in years.
- **Menopausal Status:** Whether the patient is pre-menopausal or post-menopausal.
- **Tumour Size:** Size of the tumour in millimetres.
- **Tumour Grade:** A categorical variable indicating the grade of the tumour (Grade 1, Grade 2, Grade 3).
- **Number of Positive Lymph Nodes:** Number of lymph nodes that are positive for cancer.
- **Progesterone Receptor Status (PR):** Presence of progesterone receptors (positive or negative).
- **Oestrogen Receptor Status (ER):** Presence of oestrogen receptors (positive or negative).
- **Hormonal Therapy:** Whether the patient received hormonal therapy (yes or no).

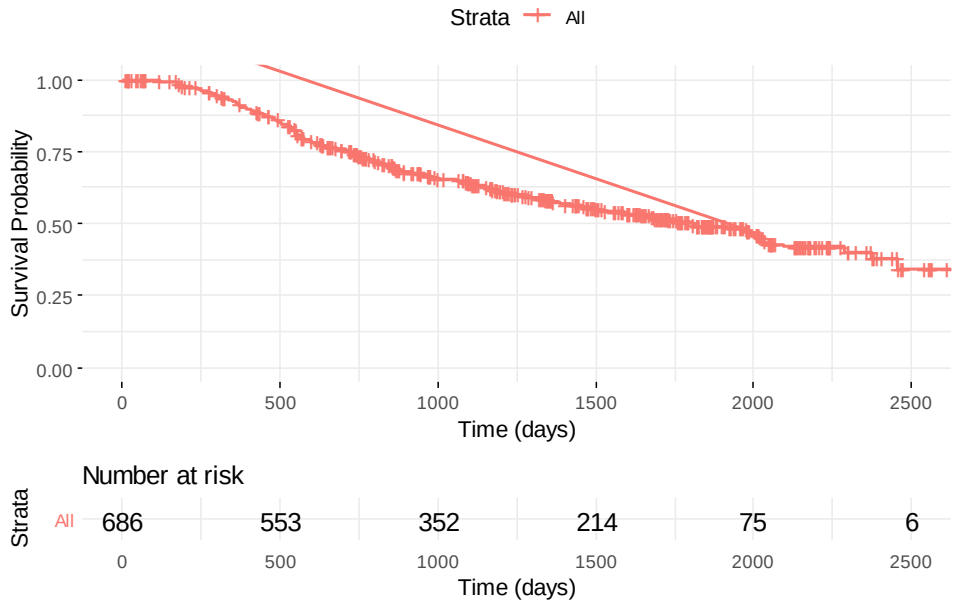
The primary goal of this analysis is to understand how these factors influence recurrence-free survival time and to identify significant predictors of survival.

Kaplan-Meier Estimation

Kaplan-Meier estimation is a non-parametric method used to estimate the survival function from lifetime data. It is particularly useful when dealing with censored data, which is common in survival analysis. The Kaplan-Meier curve provides an estimate of the survival probability over time, with confidence intervals that give an indication of the uncertainty around the estimate.

1. Overall Survival Probability

The first Kaplan-Meier curve presents the overall survival probability for the entire dataset, without stratification by any factors.



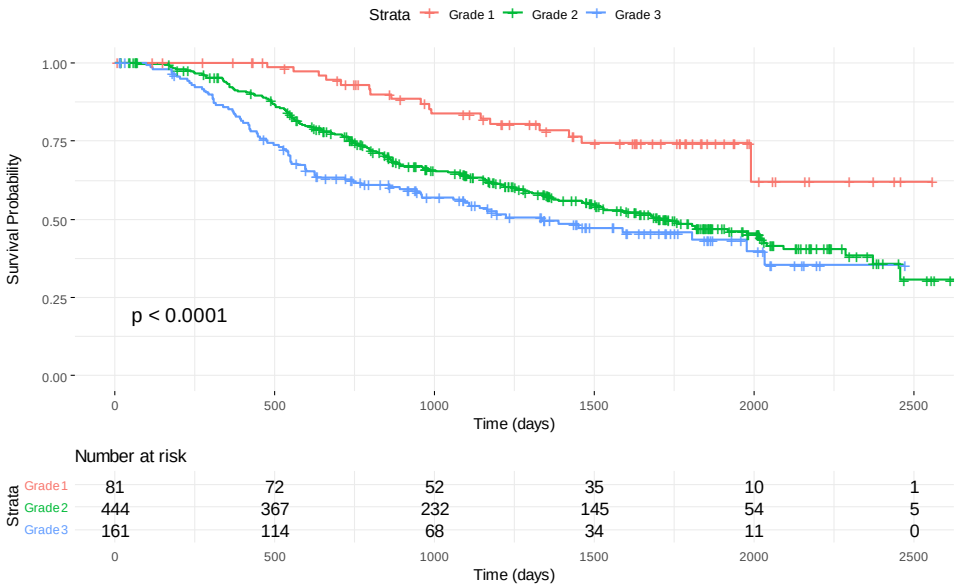
- **Number at Risk:** The number of patients at risk at various time points is displayed below the curve. The numbers decrease over time as patients experience events (recurrence) or are censored.

Interpretation:

The overall survival probability declines steadily over time, with a significant drop after around 1000 days. The curve shows that approximately 50% of the patients are still recurrence-free after about 1500 days.

2. Survival Stratified by Tumour Grade

Tumour grade is a measure of how much the cancer cells differ from healthy cells and is often associated with the aggressiveness of the tumour. The Kaplan-Meier curves are stratified by tumour grade (Grade 1, Grade 2, and Grade 3).



- **Grade 1 (Red Curve):** Represents the lowest grade, indicating well-differentiated (less aggressive) tumours.
- **Grade 2 (Green Curve):** Intermediate grade, indicating moderately differentiated tumours.
- **Grade 3 (Blue Curve):** Represents the highest grade, indicating poorly differentiated (more aggressive) tumours.

Number at Risk:

- **Grade 1:** 81, 72, 52, 35, 10, 1 at 0, 500, 1000, 1500, 2000, and 2500 days, respectively.
- **Grade 2:** 444, 367, 232, 145, 54, 5 at 0, 500, 1000, 1500, 2000, and 2500 days, respectively.
- **Grade 3:** 161, 114, 68, 34, 11, 0 at 0, 500, 1000, 1500, 2000, and 2500 days, respectively.

P-value: < 0.0001

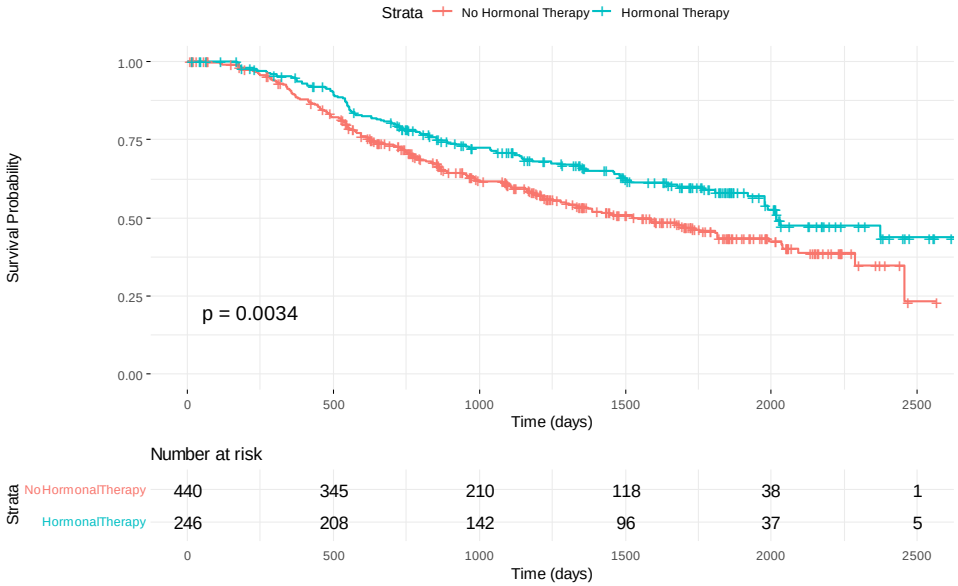
Interpretation:

The survival probability is significantly different across the tumour grades. Patients with Grade 1 tumours have the highest survival probability, followed by those with Grade 2 and

Grade 3 tumours. The sharp decline in the survival curve for Grade 3 tumours indicates a more aggressive disease with a poorer prognosis. The log-rank test confirms that the differences between the curves are statistically significant.

3. Survival Stratified by Hormonal Therapy

Hormonal therapy is a common treatment for hormone receptor-positive breast cancer. The Kaplan-Meier curves are stratified by whether or not the patient received hormonal therapy.



- **Hormonal Therapy (Blue Curve):** Represents patients who received hormonal therapy.
- **No Hormonal Therapy (Red Curve):** Represents patients who did not receive hormonal therapy.

Number at Risk:

- **No Hormonal Therapy:** 440, 345, 210, 118, 38, 1 at 0, 500, 1000, 1500, 2000, and 2500 days, respectively.

- **Hormonal Therapy:** 246, 208, 142, 96, 37, 5 at 0, 500, 1000, 1500, 2000, and 2500 days, respectively.

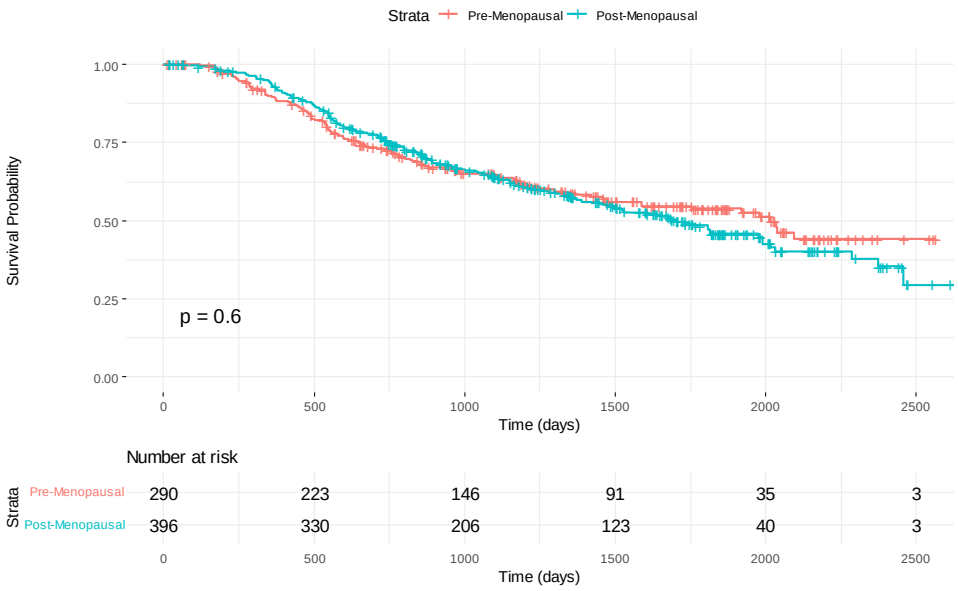
P-value: 0.0034

Interpretation:

Patients who received hormonal therapy show a higher survival probability compared to those who did not. The difference between the curves is statistically significant, indicating that hormonal therapy is effective in improving recurrence-free survival in breast cancer patients. This finding aligns with clinical practice, where hormonal therapy is known to reduce the risk of recurrence in hormone receptor-positive breast cancer.

4. Survival Stratified by Menopausal Status

Menopausal status is another important factor in breast cancer prognosis. The Kaplan-Meier curves are stratified by menopausal status (pre-menopausal and post-menopausal).



Pre-menopausal (Red Curve):

- Number at risk: 290, 223, 146, 91, 35, 3 at 0, 500, 1000, 1500, 2000, 2500 days.

Post-menopausal (Blue Curve):

- Number at risk: 396, 330, 206, 123, 40, 3 at 0, 500, 1000, 1500, 2000, 2500 days.

P-value: 0.6

Interpretation:

The Kaplan-Meier survival curves for pre-menopausal and post-menopausal women are relatively close, with no statistically significant difference observed (p=0.6). This indicates that, in this dataset, menopausal status does not significantly impact the recurrence-free survival probability over time. The survival curves are almost overlapping, suggesting that menopausal status alone may not be a strong predictor of recurrence-free survival.

Model Fitting and Further Analysis

In addition to the Kaplan-Meier estimation, further analysis can be conducted using parametric models such as exponential, Weibull, or Cox proportional hazards models. These models allow us to quantify the effect of multiple covariates on survival time and to adjust for confounding factors.

- Cox Proportional Hazards Model: This model is commonly used in survival analysis to explore the relationship between the survival time of patients and one or more predictor variables. It assumes that the hazard ratios are constant over time.
- Weibull Model: A parametric model that can accommodate increasing, constant, or decreasing hazard rates. It is flexible and can model various types of survival data.

Further model fitting and diagnostics would help refine the understanding of how these factors interact and contribute to recurrence-free survival.

Conclusion

This survival analysis provides valuable insights into the factors that influence recurrence-free survival in breast cancer patients. The key findings include:

1. Tumour Grade: A significant predictor of survival, with higher grades associated with poorer survival outcomes.

2. Hormonal Therapy: Effective in improving survival, with a statistically significant benefit observed in patients who received therapy.

3. Menopausal Status: Not a significant predictor of survival in this dataset, suggesting that other factors may play a more critical role in determining outcomes.

These findings are consistent with clinical expectations and underscore the importance of individualized treatment planning based on tumour characteristics and therapy options. Further model fitting and analysis could provide additional insights and help identify other important prognostic factors.

This report serves as a foundation for more advanced survival analysis and can be extended to explore other covariates or interactions between variables. The ultimate goal is to improve patient outcomes through better understanding and application of survival analysis in clinical practice.

References:

1. Biswas, S. (2007). Applied Stochastic Processes: A Biostatistical and Population Oriented Approach, Reprinted 2nd Ed., New Central Book Agency.

2. Elandt-Johnson R.C (1971). Probability model and Statistical Methods in Genetics, John Wiley & Sons.

3. Indrayan, A. (2008). Medical Biostatistics, 2nd Ed., Chapman and Hall/CRC.

4. Lee, E.T. and Wang, J.W. (2003). Statistical Methods for Survival data Analysis, 3rd Ed., John Wiley & Sons.

5. Narayan P. (1999). Statistical Genetics, New Age International Pvt. Ltd.

6. Miller, R. G. (2011). Survival Analysis. John Wiley & Sons.