# CLINICAL SURVIVAL ANALYSIS REPORT

**TITLE:**
SURVIVAL ANALYSIS IN CLINICAL TRIALS: KAPLAN-MEIER AND LOG-RANK TEST FINDINGS

**AUTHOR:** ABDULBASIT ADEFUGABI

**DATE:** APRIL 2025

# 1. Introduction

This project conducts a survival analysis on real-world clinical patient data using a small discovery cohort. The aim is to understand patient survival outcomes over time and explore variations across demographics and clinical characteristics using Kaplan-Meier estimation and the Log-Rank test.

# 2. Dataset Summary

**Cohort Used:** Discovery cohort (30 patients)

**Key Data Columns:**

- Patient demographics: sex, race
- Clinical features: Stage
- Survival outcomes: Dead or Alive, Time, Event
- Dates: Specimen date, Date of death, Date of last follow-up

**Cleaning Performed:**

- Standardized column names
- Converted date columns to datetime format
- Cleaned and standardized categorical values
- Checked for inconsistencies and logical errors (e.g., death before specimen date)

# 3. Methodology

**Tools:** Python, Pandas, Lifelines, Matplotlib, Seaborn

**Steps:**

1. Data preprocessing: Ensured clean date formats, categorical values, and proper structure
2. Exploratory Data Analysis: Visualized survival patterns, tumor stages, and sex differences
3. Kaplan-Meier Estimator: Estimated survival probabilities over time for males and females
4. Log-Rank Test: Compared survival curves to assess statistical significance

# 4. Results

### Kaplan-Meier Curves:

- The survival probability over time differed between males and females.
- Female patients showed slightly better survival probability than males.

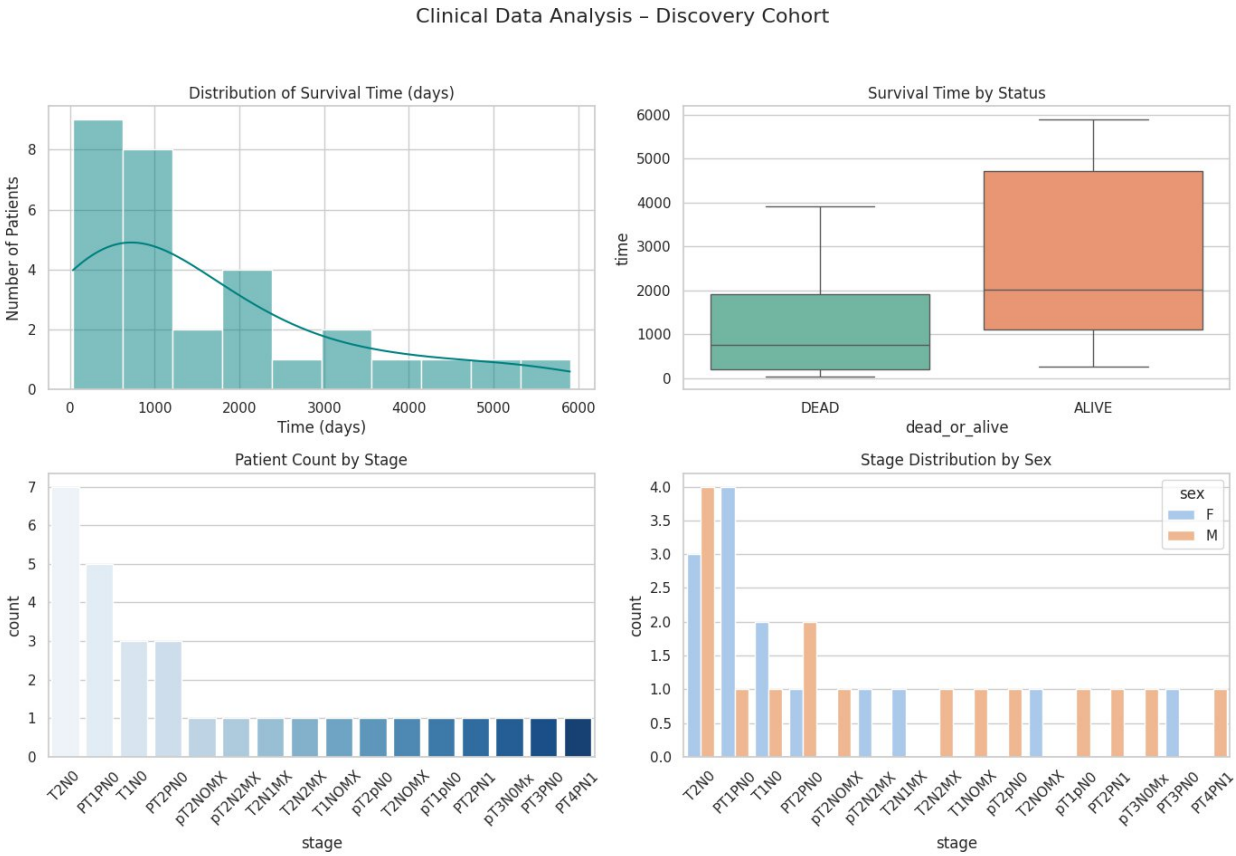
### Log-Rank Test:

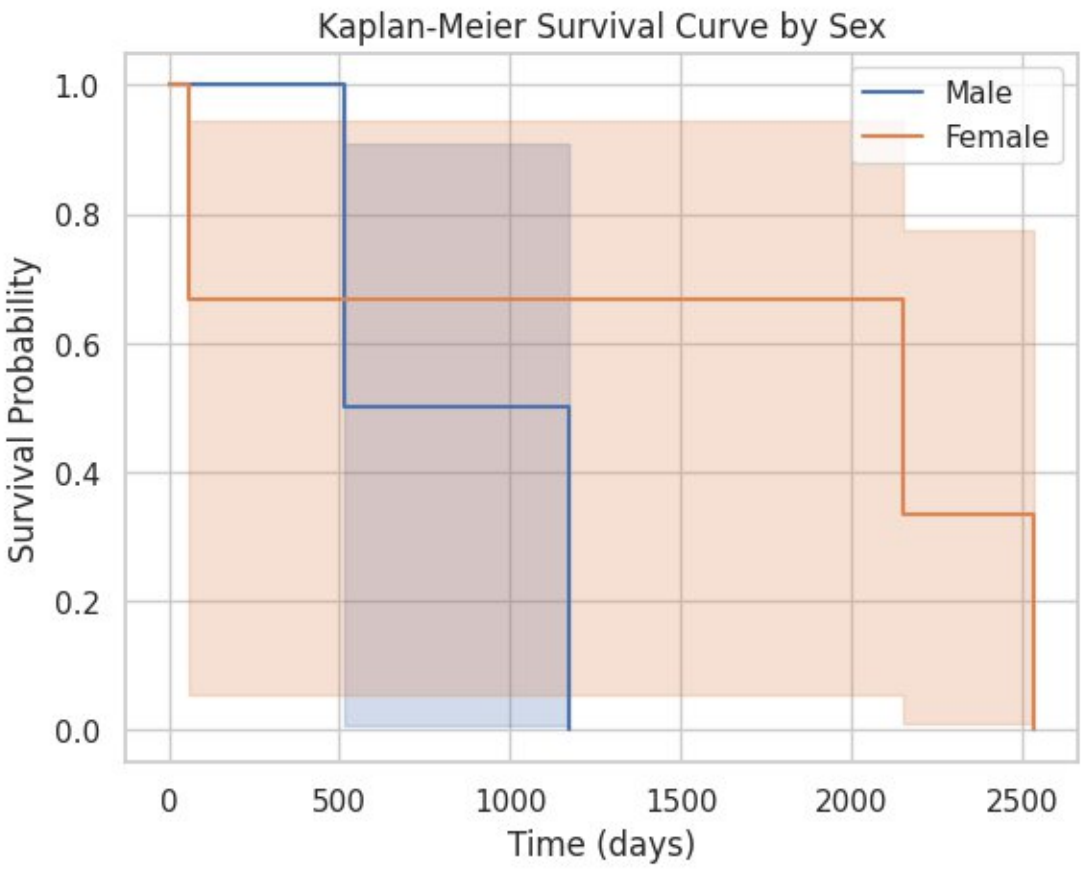
- **p-value:** 0.364 (approx.)
- **Interpretation:** No statistically significant difference in survival between male and female patients ($p > 0.05$)

### Visual Summaries:

- **Histogram:** Most patients survived 1,000–3,000 days

- **Boxplot:** Alive patients had a longer median survival than deceased ones
- **Bar Chart:** Most common stages were T2N2MX, T2N1MX, and pT2N2MX
- **Grouped Bar Chart:** Tumor stage distribution was balanced across sexes






# 5. Conclusion

This clinical data analysis uncovered key survival trends, especially across gender and tumor stage. While females demonstrated higher survival probabilities, the difference was not statistically significant based on the Log-Rank test. This method provides a powerful approach to understanding real-world patient outcomes.

# 6. Tools Used

- **Python Libraries:** Pandas, Seaborn, Matplotlib, Lifelines
- **Notebook:** Goggle Colab

- **Initial Review:** Excel

# 7. Future Scope

- Develop a Cox Proportional Hazards Model for multivariable risk prediction
- Build an interactive clinical dashboard using Streamlit or Dash
- Collaborate with medical professionals for deeper clinical interpretation