Survival Analysis of TCGA-LAML Clinical Data

Banan Bashar Aljarrah

2025-07-03

Contents

1 Abstract

This report presents a survival analysis of patients with Acute Myeloid Leukemia (AML) using clinical data from the TCGA-LAML project. The dataset was pre-processed to ensure the reliability of demographic and survival-related variables, including age, gender, and time-to-event information. Both Kaplan-Meier estimators and Cox proportional hazards regression models were applied to explore survival differences and evaluate the impact of key variables. Results suggest that patients under 60 years of age and female patients tend to have better survival outcomes. These findings are consistent with previous studies and highlight the prognostic relevance of demographic factors in AML. This analysis provides a foundation for more advanced modeling and personalized risk stratification in future research.

2 Introduction

This report provides a professional survival analysis of the TCGA-LAML dataset using R. It includes steps for data cleaning, missing value handling, exploratory data analysis, and Kaplan-Meier survival estimation by age and gender. Given the growing global burden and increasing incidence of cancer over the years, studying cancer-related outcomes such as survival is of critical importance. This analysis aims to contribute to the understanding of survival patterns in Acute Myeloid Leukemia (AML) patients by utilizing high-quality clinical data from TCGA.

3 Data Import and Cleaning

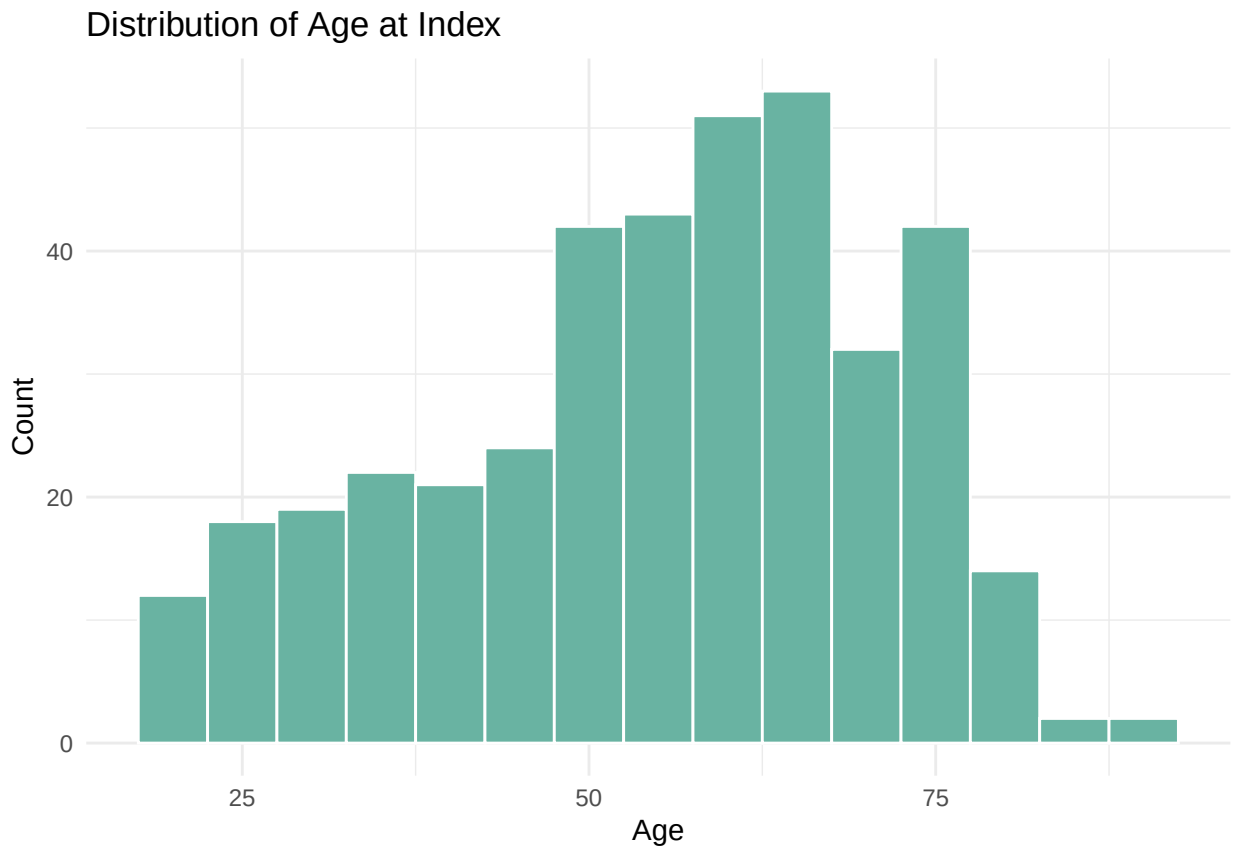
We imported the TCGA-LAML clinical dataset and performed essential data cleaning steps to ensure accuracy and consistency. Invalid or extreme values were removed (e.g., unrealistic ages), gender and vital status variables were standardized and converted to categorical data, and missing values were carefully handled. The cleaned data was then used for exploratory analysis and survival modeling.

```
## # A tibble: 6 x 210
##   project.project_id cases.case_id      cases.consent_type cases.days_to_consent
##   <chr>              <chr>              <chr>              <chr>
## 1 TCGA-LAML          008ddf20-f7fb-467~ '--                '--
## 2 TCGA-LAML          008ddf20-f7fb-467~ '--                '--
## 3 TCGA-LAML          02634746-93b2-4cc~ '--                '--
## 4 TCGA-LAML          02634746-93b2-4cc~ '--                '--
## 5 TCGA-LAML          02e4f2da-9977-425~ '--                '--
## 6 TCGA-LAML          02e4f2da-9977-425~ '--                '--
## # i 206 more variables: cases.days_to_lost_to_followup <chr>,
## #   cases.disease_type <chr>, cases.index_date <chr>,
## #   cases.lost_to_followup <chr>, cases.primary_site <chr>,
## #   cases.submitter_id <chr>, demographic.age_at_index <dbl>,
## #   demographic.age_is_obfuscated <lgl>, demographic.cause_of_death <chr>,
## #   demographic.cause_of_death_source <chr>,
## #   demographic.country_of_birth <chr>, ...
```

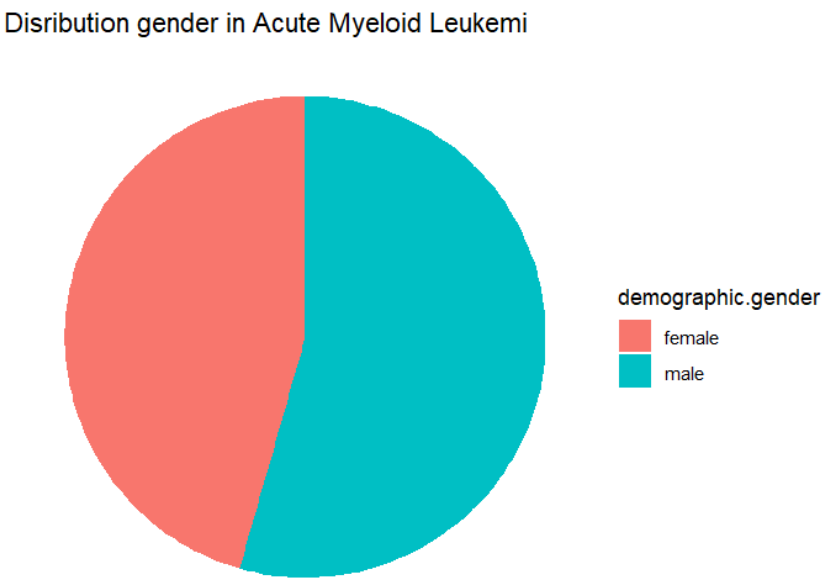
3.1 Handle Missing and Erroneous Values

4 Exploratory Data Analysis (EDA)

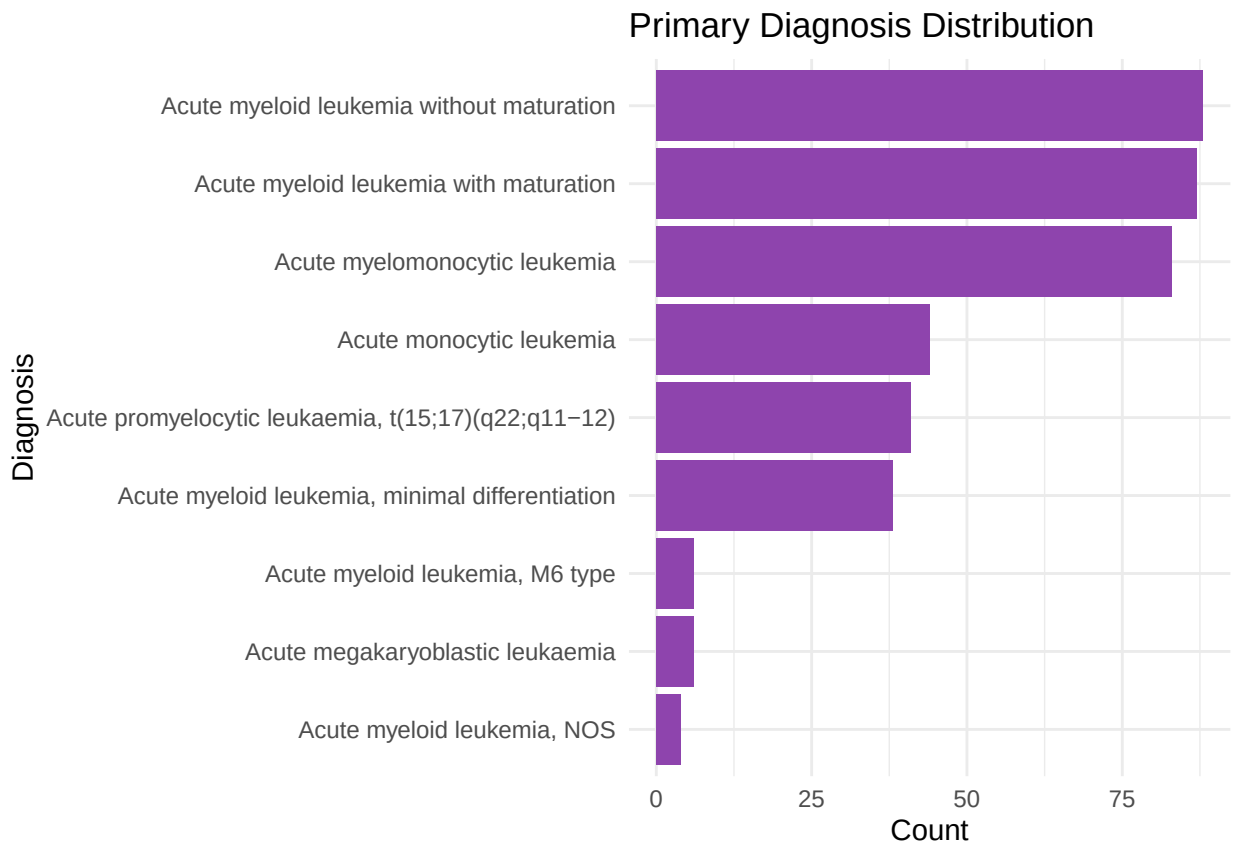
4.1 Age at Index Distribution



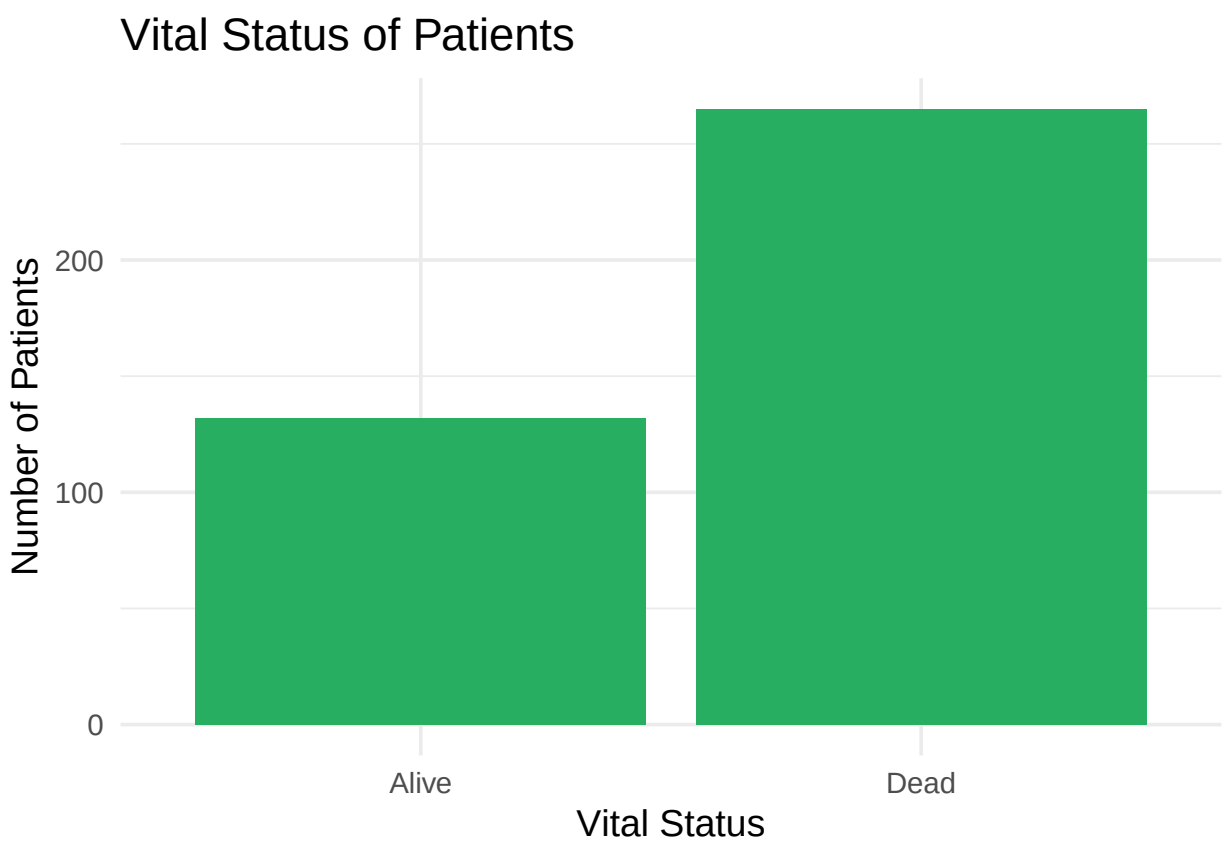
4.2 Gender Distribution



4.3 Primary Diagnosis Distribution

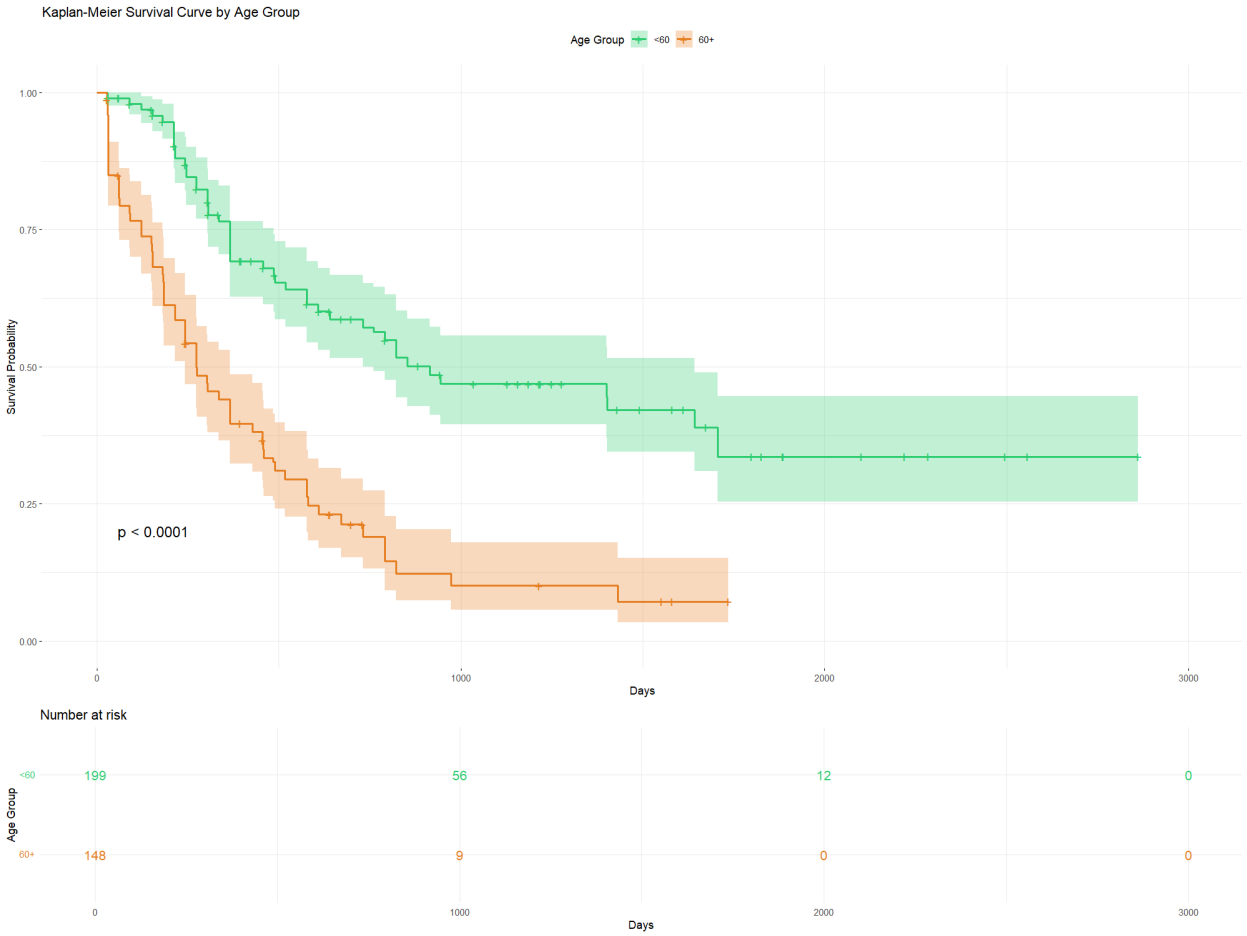


4.4 Distribution of patients by Vital status

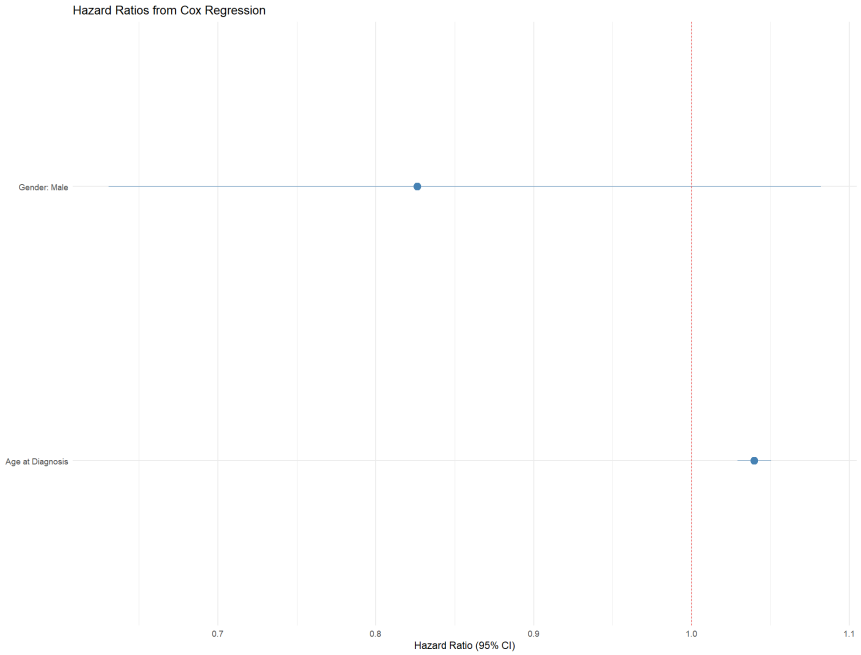


5 Survival Analysis

5.1 Kaplan-Meier Curve by Gender



5.2 Cox Regression: Impact of Age and Gender on Survival in AML Patients



---

6 Scientific Interpretation

Gender:
The Kaplan-Meier survival curves indicate that female patients have a slightly longer overall survival compared to male patients. This trend may reflect underlying biological or hormonal differences, treatment responsiveness, or supportive care variations. However, further studies are required to establish causality.

Age:
Patients under 60 years of age demonstrated significantly improved survival outcomes relative to older patients. This difference may be explained by better treatment tolerance, fewer comorbidities, and a higher likelihood of receiving intensive or curative therapies in younger populations.

Cox Regression:
The Cox proportional hazards model confirmed that both gender and age are independently associated with survival outcomes in AML. Specifically, being younger than 60 was significantly protective, while the effect of gender, although present, was less pronounced.

---

7 Summary

This survival analysis of TCGA-LAML clinical data revealed distinct survival patterns by age and gender. Younger patients (<60 years) showed markedly better outcomes, and females had a slight survival advantage compared to males. These findings align with established literature and provide insights into risk stratification.

The use of both Kaplan-Meier curves and Cox regression modeling offered a comprehensive view of survival dynamics in AML. The observed patterns may guide clinicians in patient counseling and contribute to more personalized treatment planning.

Future research could benefit from multivariable modeling including genetic, molecular, and treatment-related variables to enhance predictive accuracy.

8 References

[1] National Cancer Institute – SEER. Acute Myeloid Leukemia Survival Statistics. Available at: https://seer.cancer.gov/statfacts/html/amyl.html

[2]: Juliusson, G., et al. (2011). Age and gender impact survival in AML: A population-based study. *Blood*, 118(18), 4591–4597. https://pubmed.ncbi.nlm.nih.gov/21725052/