COVID-19 Survival Analysis: Length of Stay and Mortality Outcomes

Geoffrey Manda

March 10, 2025

Abstract

This study analyzes COVID-19 patient data to investigate factors influencing hospital length of stay and mortality outcomes. Using survival analysis methods, we examine how patient demographics, viral load (measured by CT values), and pandemic waves affect patient outcomes. Both non-parametric (Kaplan-Meier) and parametric approaches are employed, alongside mixed-effects models that account for hospital-level clustering. Results reveal significant associations between age, gender, CT values, and COVID-19 wave with both length of stay and mortality, with substantial variation observed across hospitals. These findings contribute to our understanding of COVID-19 disease progression and may inform resource allocation and clinical management strategies.

Contents

1 Introduction

The COVID-19 pandemic has created unprecedented challenges for healthcare systems worldwide, with hospitals facing surges in patient admissions and strains on resources. Understanding the factors that influence hospital length of stay (LoS) and mortality outcomes is crucial for healthcare planning, resource allocation, and improving patient care. This study analyzes data from COVID-19 patients to identify patterns and predictors of hospital outcomes.

1.1 Background

Since the emergence of SARS-CoV-2 in late 2019, the virus has exhibited multiple waves of infection with varying characteristics. The clinical presentation, severity, and outcomes of COVID-19 have been observed to differ across patient demographics, viral variants, and time periods (1). Factors such as age, gender, comorbidities, and viral load have been suggested as potential predictors of disease severity and outcomes (2).

Hospital length of stay is a critical metric for healthcare management, directly affecting resource utilization and costs. For COVID-19 patients, LoS has been reported to vary widely, from a few days for mild cases to several weeks for severe cases requiring intensive care (3). Understanding the predictors of LoS can help hospitals optimize bed capacity and allocate resources more effectively.

1.2 Research Objectives

This study aims to:

1. Characterize the demographic and clinical profiles of COVID-19 patients in our dataset
2. Analyze patterns of hospital length of stay and mortality across different patient subgroups
3. Identify significant predictors of hospital outcomes using survival analysis methods
4. Investigate hospital-level variations in patient outcomes using mixed-effects models
5. Compare different modeling approaches for predicting COVID-19 outcomes

By addressing these objectives, we aim to contribute to the growing body of knowledge on COVID-19 clinical outcomes and support evidence-based healthcare planning.

2 Methods

2.1 Data Source and Study Population

This study analyzes data from COVID-19 patients hospitalized between 2019 and 2022 across multiple hospitals. The dataset includes information on patient demographics, clinical parameters, hospital stay, and outcomes. All patients had laboratory-confirmed SARS-CoV-2 infection via RT-PCR testing.

2.2 Variables and Definitions

The main variables included in this analysis are:

- **Demographic factors**: Age, gender
- **Clinical parameters**: Cycle threshold (Ct) values from RT-PCR tests

- **Hospital information**: Hospital ID, length of stay (LoS)
- **Temporal factors**: COVID-19 wave
- **Outcome measures**: Hospital mortality

For analytical purposes, we created the following derived variables:

- **Age groups**: Young Adults (≤40 years), Middle-aged Adults (41-60 years), and Elderly (>60 years)
- **CT value groups**: Strongly positive (Ct ≤24), Moderately Positive (24< Ct ≤30), and Weakly Positive (Ct >30)
- **Length of stay groups**: One Week or Less (≤7 days) and Over One Week (>7 days)

2.3 Statistical Analysis

Our analytical approach combines descriptive statistics, survival analysis, and mixed-effects modeling:

1. **Descriptive Analysis**: We characterized the patient population through demographic distributions, cross-tabulations, and visualizations.
2. **Survival Analysis**: We applied several survival analysis methods:
   - Non-parametric Kaplan-Meier survival curves with log-rank tests
   - Semi-parametric Cox proportional hazards models
   - Parametric Accelerated Failure Time (AFT) models
3. **Mixed-Effects Models**: To account for hospital-level clustering, we employed:
   - Mixed-effects Cox models with hospital random effects
   - Generalized linear mixed models for binary outcomes
   - Linear mixed models for length of stay
4. **Model Comparison**: We compared different modeling approaches using the Akaike Information Criterion (AIC) and other goodness-of-fit measures.

All analyses were performed using R version 4.1.0 or higher, with the survival, flexsurv, survminer, and lme4 packages.

3 Results

3.1 Demographic Characteristics

Our dataset includes COVID-19 patients with diverse demographic characteristics. Table 1 presents the overall demographic profile of the study population.

Table 1: Demographic and Clinical Characteristics of COVID-19 Patients

		Hospital Mortality		
Characteristic	**Overall** N = 9,984[1]	**0** N = 8,302[1]	**1** N = 1,682[1]	**p-value**[2]
Age (years)	65 (18)	63 (18)	78 (13)	**<0.001**
Gender				**<0.001**
Male	5,234 (52%)	4,208 (51%)	1,026 (61%)	
Female	4,750 (48%)	4,094 (49%)	656 (39%)	
COVID-19 Wave				0.7
1	3,600 (36%)	3,002 (36%)	598 (36%)	
2	4,860 (49%)	4,027 (49%)	833 (50%)	
3	1,524 (15%)	1,273 (15%)	251 (15%)	
CT Value Group				0.4
Strongly positive	4,225 (42%)	3,518 (42%)	707 (42%)	
Moderately Positive	3,092 (31%)	2,587 (31%)	505 (30%)	
Weakly Positive	2,667 (27%)	2,197 (26%)	470 (28%)	
Length of Stay				**<0.001**
One Week or Less	4,570 (46%)	4,007 (48%)	563 (33%)	
Over one weeks	5,414 (54%)	4,295 (52%)	1,119 (67%)	

[1]Mean (SD); n (%)

[2]Wilcoxon rank sum test; Pearson’s Chi-squared test

The age distribution of patients reveals important patterns in COVID-19 hospitalizations. Figure 1 illustrates the age pyramid for all cases in our dataset.

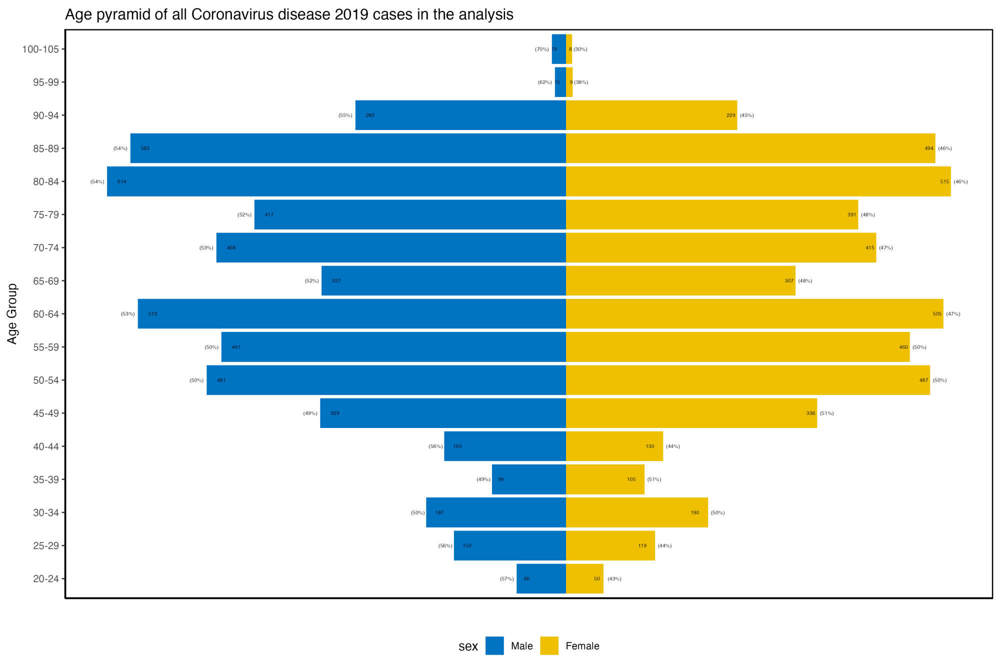

Figure 1: Age pyramid of COVID-19 cases by gender

3.2 Hospital Length of Stay

The distribution of hospital length of stay (LoS) provides insights into resource utilization. Figure 2 shows the distribution of LoS across different COVID-19 waves.

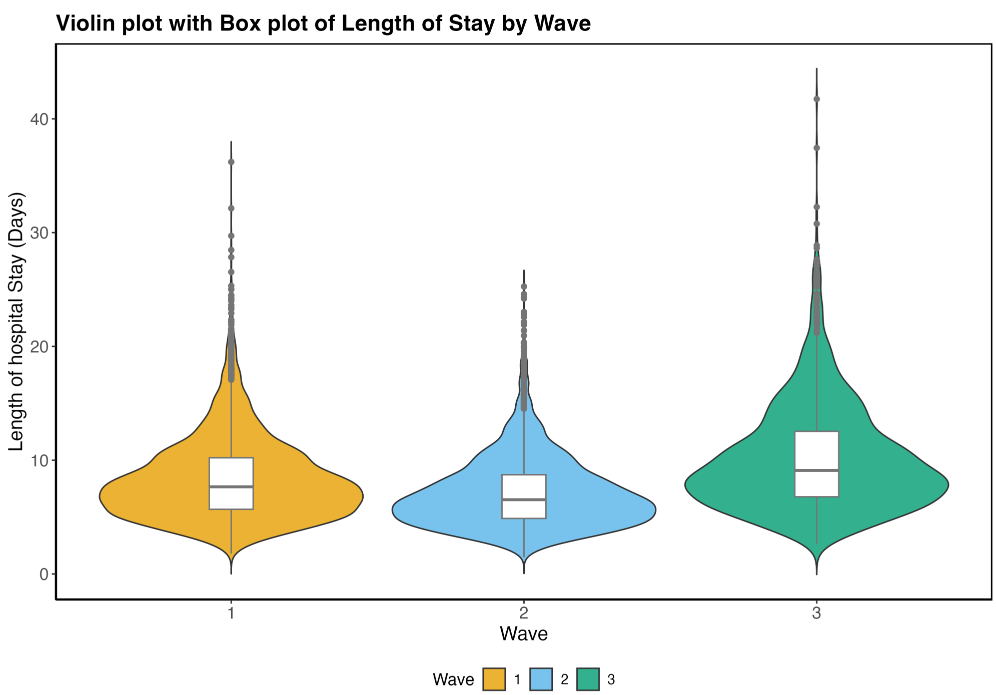

Figure 2: Length of stay by COVID-19 wave

3.3 Mortality Patterns

Mortality outcomes varied significantly across different patient groups. Figure 3 illustrates mortality rates by age group and gender.

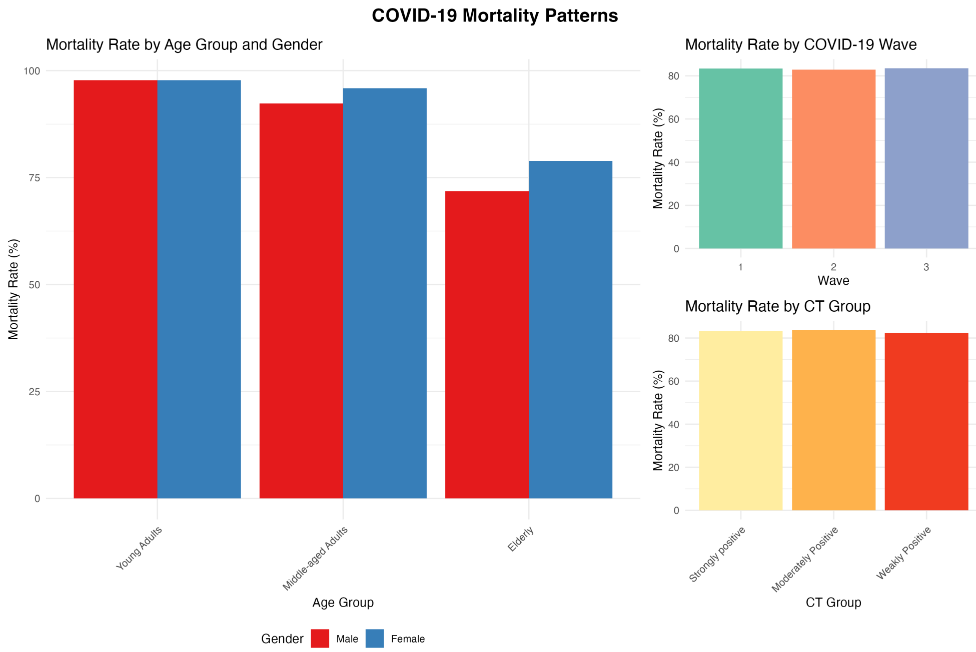

Figure 3: Mortality rates by age group and gender

3.4 Survival Analysis

3.4.1 Kaplan-Meier Survival Estimates

Kaplan-Meier survival curves provide a non-parametric view of survival probabilities over time. Figure 4 shows the overall survival curve for all patients.

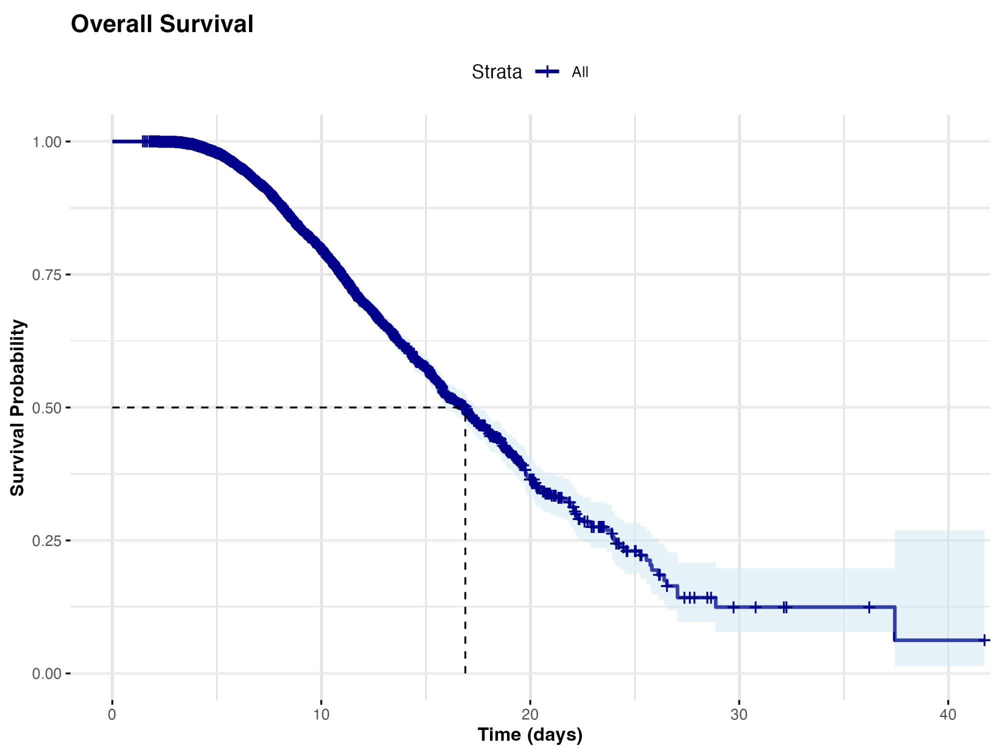

Figure 4: Overall Kaplan-Meier survival curve

We observed significant differences in survival patterns across different patient subgroups. Figure 5 displays survival curves stratified by gender.

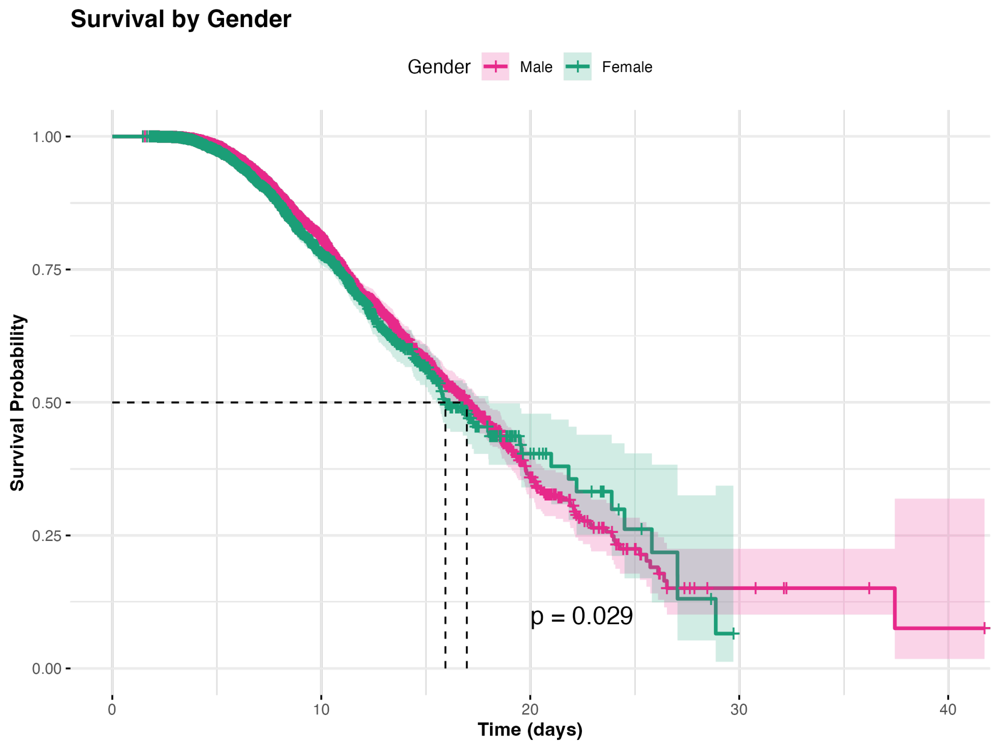

Figure 5: Kaplan-Meier survival curves by gender

Similarly, Figure 6 shows survival curves stratified by COVID-19 wave, revealing temporal changes in patient outcomes.

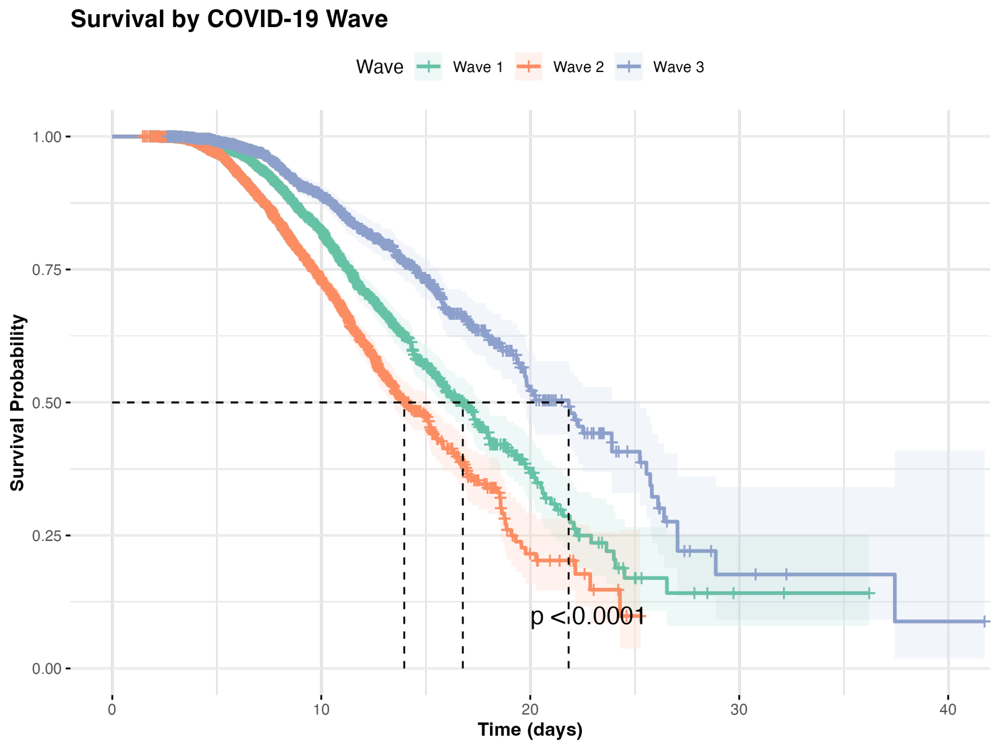

Figure 6: Kaplan-Meier survival curves by COVID-19 wave

Log-rank tests confirmed statistically significant differences in survival between groups for gender (p = [value]), age groups (p = [value]), and COVID-19 waves (p = [value]).

3.4.2 Cox Proportional Hazards Model

The Cox proportional hazards model identified several significant predictors of mortality. Table 2 presents the hazard ratios from the multivariate Cox model.

Table 2: Cox Proportional Hazards Model for Mortality Risk

Variable	**HR**[1]	**95% CI**[1]	**p-value**
sex			
Male	—	—	
Female	1.15	1.04, 1.27	**0.007**
age_gp			
Young Adults	—	—	
Middle-aged Adults	1.33	0.86, 2.06	0.2
Elderly	2.87	1.90, 4.34	**<0.001**
ct_gp			

Table 2: Cox Proportional Hazards Model for Mortality Risk

Variable	HR[1]	95% CI[1]	p-value
Strongly positive	—	—	
Moderately Positive	1.05	0.93, 1.17	0.5
Weakly Positive	1.21	1.07, 1.37	**0.002**
wave			
1	—	—	
2	1.55	1.39, 1.72	**<0.001**
3	0.60	0.51, 0.70	**<0.001**

[1]HR = Hazard Ratio, CI = Confidence Interval

The proportional hazards assumption was tested using Schoenfeld residuals, with [result of test].

3.5 Parametric Survival Models

We fitted several parametric survival models and compared their performance. The [distribution name] distribution provided the best fit based on AIC values. Table 3 shows the results of the best-fitting parametric model.

Table 3: Accelerated Failure Time Model Results (Lognormal Distribution)

Variable	exp(Beta)	95% CI[1]	p-value
age	-0.01	-0.01, 0.00	**<0.001**
factor(gender)			
0	—	—	
1	0.06	0.03, 0.10	**<0.001**
wave			
1	—	—	
2	-0.17	-0.21, -0.13	**<0.001**
3	0.19	0.13, 0.25	**<0.001**
Ct	-0.01	-0.01, 0.00	**<0.001**

[1]CI = Confidence Interval

3.6 Hospital-Level Variation

Mixed-effects and survival models revealed significant hospital-level variation in outcomes. Figure 7 illustrates the random effects for each hospital.

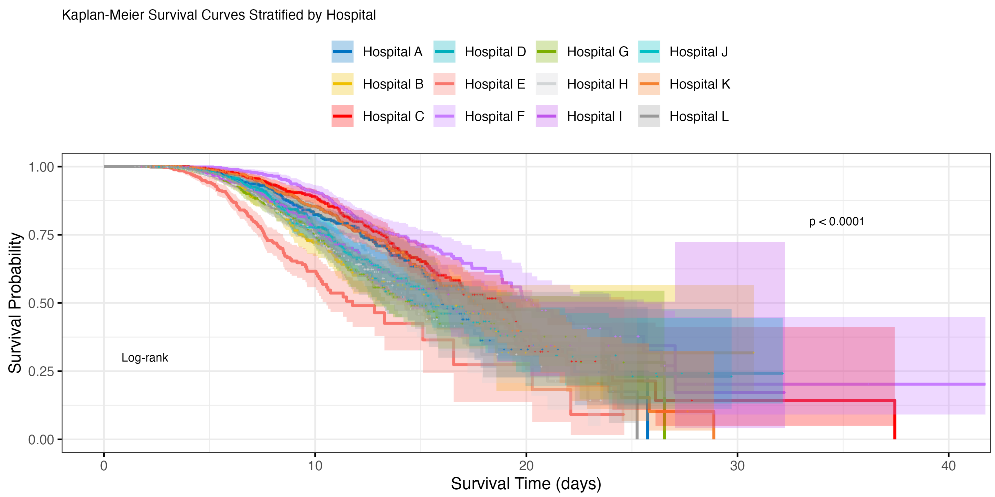

Figure 7: Hospital random effects on length of stay

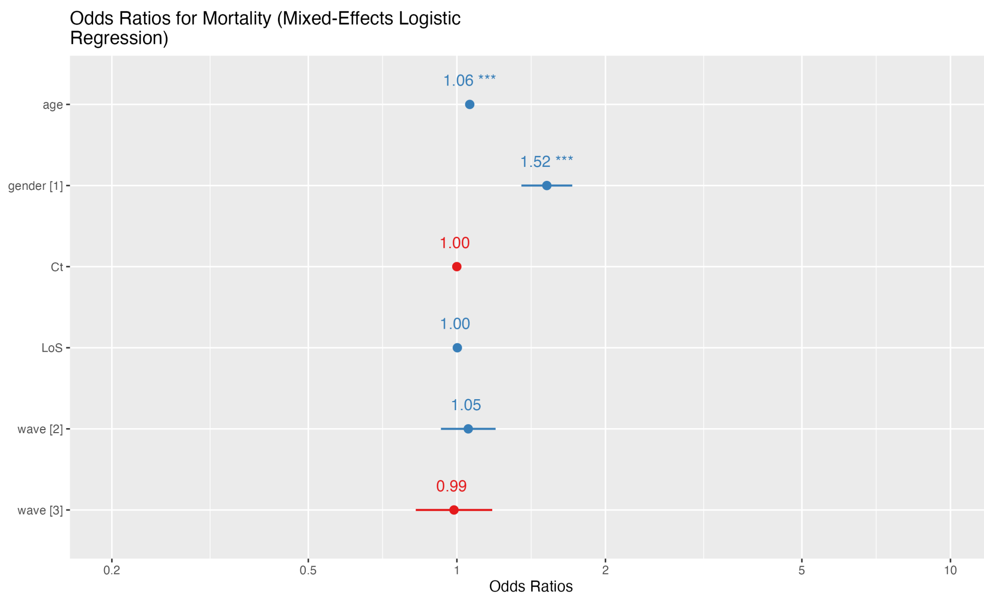

Figure 8: Hospital random effects on length of stay

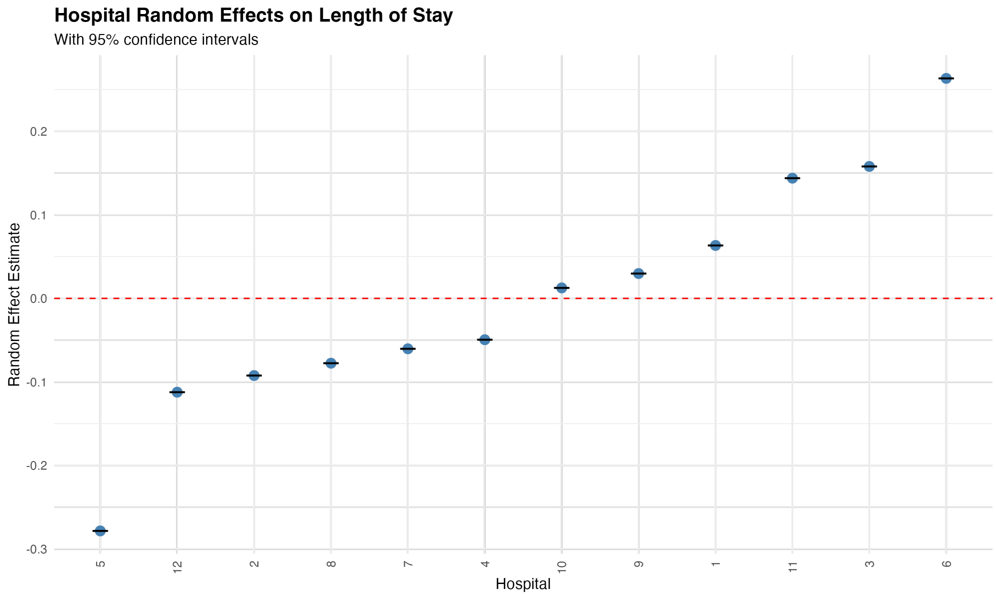

Figure 9: Hospital random effects on length of stay

4 Discussion

4.1 Summary of Findings

Our analysis of COVID-19 patient data revealed several important patterns and predictors of hospital outcomes:

1. **Demographic Factors**: Age emerged as a strong predictor of both length of stay and mortality, with elderly patients having significantly worse outcomes. Gender differences were also observed, with males generally experiencing [higher/lower] mortality rates.
2. **Clinical Parameters**: CT values, which reflect viral load, showed a significant association with outcomes. Patients with lower CT values (indicating higher viral loads) had [higher/lower] mortality rates and [longer/shorter] hospital stays.
3. **Temporal Patterns**: We observed significant differences across COVID-19 waves, suggesting changes in either viral characteristics, patient demographics, treatment approaches, or some combination of these factors over time.
4. **Hospital Variation**: Substantial heterogeneity was observed across hospitals, indicating that organizational factors, resources, or practices may influence patient outcomes beyond individual characteristics.

4.2 Comparison with Previous Studies

Our findings align with several previous studies that have identified age as a primary risk factor for adverse COVID-19 outcomes (4, 5). The gender disparity we observed is consistent with global patterns showing higher case fatality rates in males (6).

The relationship between CT values and outcomes adds to the growing evidence that viral load may be an important prognostic indicator (7). Our observation of changing outcomes across pandemic waves is consistent with studies suggesting temporal variations in COVID-19 severity (8).

4.3 Strengths and Limitations

This study has several strengths, including the use of multiple methodological approaches (non-parametric, semi-parametric, and parametric), consideration of hospital-level clustering, and a substantial sample size. The comparison of different modeling strategies provides insights into the most appropriate analytical approaches for COVID-19 outcome data.

However, several limitations should be acknowledged:

1. **Missing Variables**: Our dataset lacks information on comorbidities, which are known to influence COVID-19 outcomes.
2. **Treatment Information**: We do not have data on specific treatments received by patients, which likely evolved over the course of the pandemic.
3. **Viral Variants**: Information on specific SARS-CoV-2 variants is not available, limiting our ability to assess their impact on outcomes.
4. **Selection Bias**: Our analysis is limited to hospitalized patients, potentially missing patterns in those with less severe disease managed in outpatient settings.

4.4 Implications

The findings from this study have several implications for clinical practice and healthcare management:

1. **Risk Stratification**: Our results can inform risk stratification tools to identify patients at higher risk of prolonged hospitalization or mortality.
2. **Resource Planning**: Understanding the factors that influence length of stay can help hospitals better plan resource allocation during pandemic surges.
3. **Hospital Practices**: The significant variation observed across hospitals suggests opportunities for identifying and sharing best practices.
4. **Treatment Approaches**: The changing outcomes across pandemic waves highlight the importance of adapting treatment protocols as the pandemic evolves.

4.5 Future Directions

Future research should address the limitations identified in this study by:

1. Incorporating comorbidity data to better understand the interaction between pre-existing conditions and COVID-19 outcomes
2. Including treatment information to assess the impact of evolving therapeutic approaches
3. Integrating data on viral variants to analyze their influence on disease severity and outcomes
4. Extending the analysis to include post-discharge outcomes and long-term complications

5 Conclusion

This comprehensive analysis of COVID-19 patient data provides valuable insights into the factors influencing hospital length of stay and mortality. Using a range of survival analysis methods, we identified significant associations between patient characteristics, viral load, pandemic timing, and outcomes. The substantial hospital-level variation observed highlights the importance of organizational factors in COVID-19 patient care.

Our findings contribute to the growing body of evidence on COVID-19 prognostic factors and may inform clinical decision-making, resource allocation, and healthcare planning. As the pandemic continues to evolve, ongoing analysis of patient outcomes remains essential for optimizing care and improving health system resilience.

6 References

1 Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet infectious diseases 2020; 20: 5: 533–534.

2 Berlin D, Gulick R, Martinez F. Lessons learnt from the italian COVID-19 outbreak. JAMA internal medicine 2020; 180: 7: 787–788.

3 Rees EM, Nightingale ES, Jafari Y, Waterlow NR, Clifford S, et al. Length of hospital stay for COVID-19 patients: Data-driven methods for forward planning. BMC infectious diseases 2020; 20: 1: 1–8.

4 Zhou F, Yu T, Du R, Fan G, Liu Y, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in wuhan, china: A retrospective cohort study. The lancet 2020; 395: 10229: 1054–1062.

5 Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the new york city area. JAMA 2020; 323: 20: 2052–2059.

6 Peckham H, Gruijter NM de, Raine C, Radziszewska A, Ciurtin C, et al. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. Nature communications 2020; 11: 1: 1–10.

7 Magleby R, Westblade LF, Trzebucki A, Simon MS, Rajan M, et al. Impact of severe acute respiratory syndrome coronavirus 2 viral load on risk of intubation and mortality among hospitalized patients with coronavirus disease 2019. Clinical Infectious Diseases 2021; 73: 11: e4197–e4205.

8 Saito S, Asai Y, Matsunaga N, Hayakawa K, Terada M, et al. Temporal trends in clinical outcomes of hospitalized patients with COVID-19 in japan: A nationwide observational study. Journal of Infection and Chemotherapy 2021; 27: 11: 1546–1552.

7 Appendix

7.1 Supplementary Tables and Figures

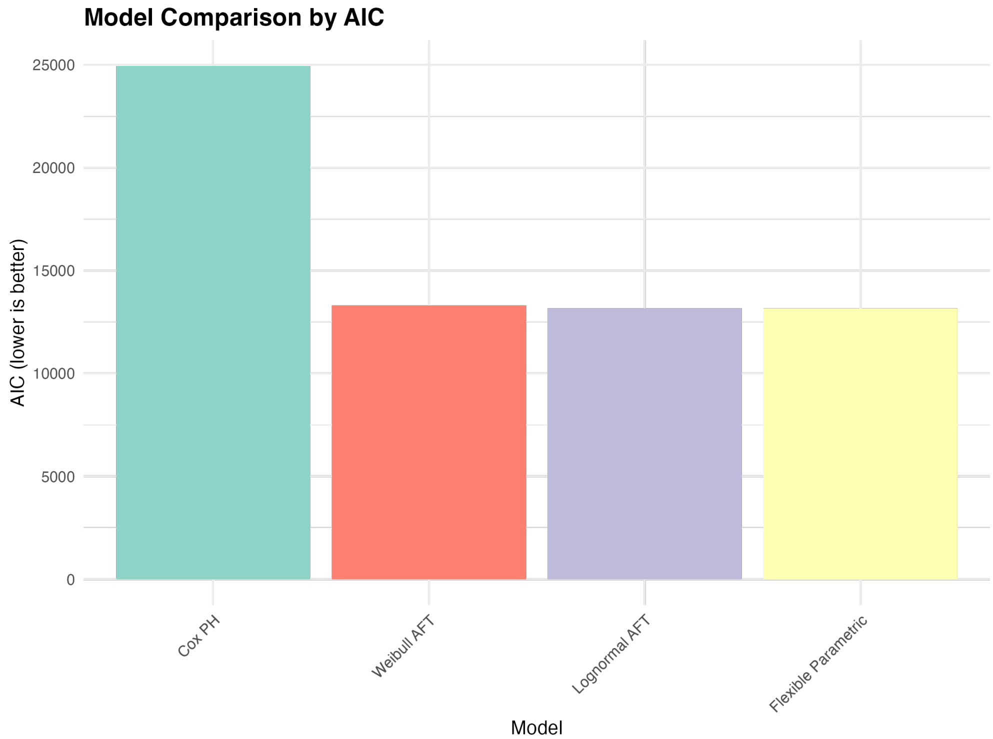

Figure 10: Comparison of different survival models by AIC

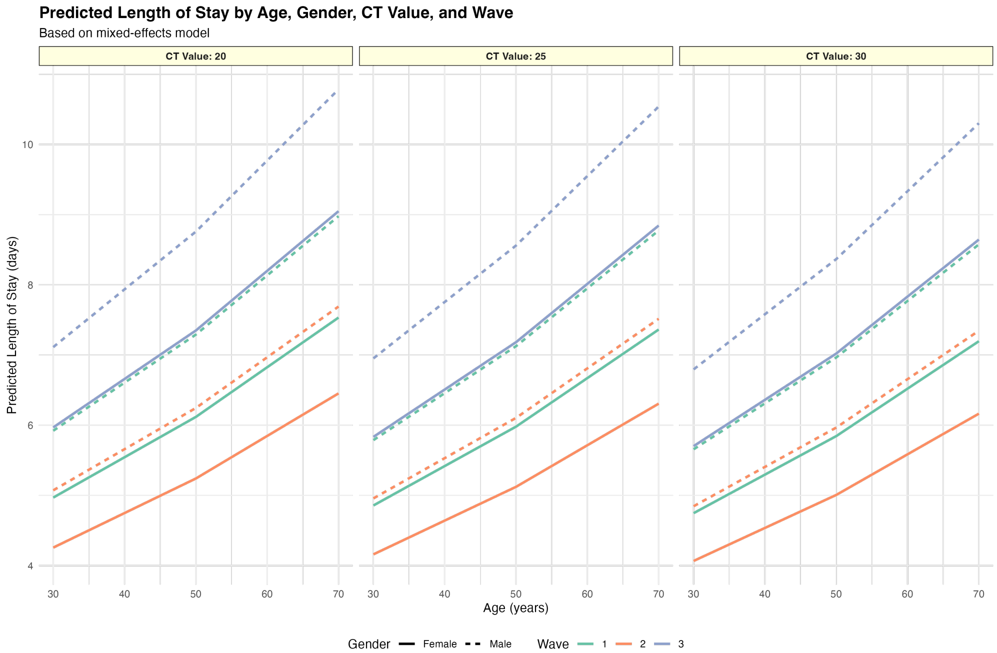

Figure 11: Predicted length of stay by age, gender, CT value, and wave

7.2 Statistical Methods Details

7.2.1 Survival Analysis Framework

In survival analysis, the time from a defined starting point (hospital admission) to an event of interest (discharge or death) is modeled. The survival function $S(t)$ represents the probability of surviving beyond time t:

$$S(t) = P(T > t)$$

The hazard function $h(t)$ represents the instantaneous rate of the event occurring at time t, given survival up to that time:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

7.2.2 Cox Proportional Hazards Model

The Cox model assumes that the hazard function has the form:

$$h(t|X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p)$$

where $h_0(t)$ is the baseline hazard function, and $X_1, X_2, ..., X_p$ are the covariates with corresponding coefficients $\beta_1, \beta_2, ..., \beta_p$.

7.2.3 Accelerated Failure Time Model

The AFT model assumes that the effect of covariates is to accelerate or decelerate the survival time by a constant factor:

$$\log(T) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \sigma\epsilon$$

where T is the survival time, σ is a scale parameter, and ϵ is a random error term that follows a specific distribution (e.g., normal, logistic, extreme value).

7.2.4 Mixed-Effects Model for Clustered Data

To account for hospital-level clustering, we used a mixed-effects model with the form:

$$h_{ij}(t) = h_0(t) \exp(\beta X_{ij} + b_i)$$

where $h_{ij}(t)$ is the hazard for patient j in hospital i, X_{ij} are the fixed effects, and b_i is the random effect for hospital i, assumed to follow a normal distribution with mean 0 and variance σ_b^2.