Innovation in Computer and Data Sciences
2(2), 29–47
DOI: https://doi.org/10.64389/icds.2026.02285
https://sphinxsp.org/journal/index.php/icds/

Research article

Comparative Assessment of Parametric Accelerated Failure Time and Cure Models for Survival Analysis in Clinical Studies

Egbodo Peter Odeh[1,*], **Chinwendu Alice Uzuke**[1]

[1] Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University, P.O. Box 5025 Awka, Nigeria

* **Correspondence:** egbodo2002@gmail.com

ARTICLE INFO

Keywords:
Survival Analysis
Accelerated Failure Time Model
Cure Model
COVID-19
Parametric Survival Models.

Mathematics Subject Classification:
62N01, 62N02

Important Dates:
Received: 4 May 2026
Revised: 4 June 2026
Accepted: 5 June 2026
Online: 28 June 2026

Copyright © 2026 by the authors. Published under Creative Commons Attribution (CC BY) license.

ABSTRACT

This study examined the survival outcomes of 322 patients diagnosed with COVID-19 and admitted to hospitals in Campinas, São Paulo, Brazil. Parametric Accelerated Failure Time (AFT) and cure models were applied to evaluate survival patterns, identify factors influencing patient survival, and compare the suitability of different parametric distributions for modeling COVID-19 survival outcomes. Model performance was assessed using log-likelihood and Akaike Information Criterion (AIC). Among the AFT models, the Weibull AFT model provided the best fit to the data (log-likelihood = -436.93, AIC = 891.85), outperforming the log-normal and exponential AFT models. Similarly, the Weibull cure model demonstrated superior performance among the cure models (log-likelihood = -441.71, AIC = 903.42). Results from the AFT models showed that age, diabetes, and neurological disorders significantly influenced survival time, while age was the only covariate consistently significant in the cure models. Other factors, including sex, asthma, heart disease, and obesity, were not statistically significant. The findings underscore the effectiveness of the Weibull distribution for modelling survival outcomes and cure fractions, while highlighting the value of cure models in providing a more comprehensive understanding of long-term patient survival.

1. Introduction

Survival analysis has emerged as an indispensable statistical framework for modelling the time until the occurrence of clinically relevant events, often referred to as “time-to-event” outcomes. In medical research,

such events may include death, disease recurrence, or recovery, all of which play central roles in evaluating treatment efficacy and guiding clinical decision-making. Unlike traditional regression approaches that assume normally distributed outcomes, survival data present unique complexities, notably censoring and non-constant hazard rates. Censoring occurs when the event of interest has not been observed for some patients by the end of the study period [15, 3, 19], leading to incomplete data that, if not properly addressed, can bias results. Survival analysis techniques are therefore uniquely suited for medical and epidemiological studies, offering robust methods for extracting meaningful insights from incomplete observations [4, 9, 14].

The importance of survival analysis became particularly evident during the COVID-19 pandemic. With the global spread of SARS-CoV-2, health systems faced overwhelming caseloads, and researchers sought to identify prognostic factors, estimate survival probabilities, and optimize the allocation of limited healthcare resources. Early studies demonstrated that demographic and clinical factors, including advanced age, male sex, and comorbidities such as diabetes, obesity, and cardiovascular disease, were strongly associated with increased risk of mortality among infected patients [13, 8, 6, 2, 7, 10, 16, 23]. Accurate prediction of survival times became essential for patient risk stratification and informed healthcare planning. These challenges highlighted the relevance of survival models capable of accommodating heterogeneous patient characteristics while capturing the dynamics of mortality risk across different stages of disease progression.

Parametric survival models, particularly Accelerated Failure Time (AFT) models, provide a powerful approach for modelling clinical survival data [21, 12]. By assuming a specific probability distribution for survival times, such as the Weibull, log-normal, or exponential distribution, parametric methods offer smooth and interpretable survival and hazard functions, efficient parameter estimation, and the ability to extrapolate beyond observed follow-up periods. Compared to non-parametric approaches like the Kaplan–Meier estimator or semi-parametric models such as the Cox proportional hazards model, parametric methods can be more statistically efficient when their assumptions hold, especially in smaller or heterogeneous samples [17]. Among these, the Weibull AFT model is widely used for its flexibility in capturing increasing, decreasing, or constant hazards [22, 5], the log-normal AFT model is suitable for scenarios where hazard rates rise to a peak before declining [11], and the exponential AFT model provides a simpler yet interpretable representation under the assumption of constant hazard rates [18].

In addition to standard parametric approaches, cure models have become increasingly relevant in clinical research. These models recognize that, for certain conditions, a subset of patients may never experience the event of interest, effectively being "cured." Mixture cure models combine parametric survival distributions with logistic regression components to estimate both survival time and the probability of cure [1]. This feature is particularly valuable in infectious disease research, including COVID-19, where some patients may achieve full recovery and remain free of long-term mortality risk related to the disease.

The present study aims to compare the assessment of Weibull, log-normal, and exponential AFT models, along with their corresponding cure model extensions, in estimating survival times among COVID-19 patients in Campinas, Brazil. The study focuses on identifying significant prognostic factors, including age, sex, and comorbidities such as heart disease, asthma, diabetes, neurological conditions, and obesity, while comparing the efficiency and explanatory power of different parametric approaches. By doing so, the analysis contributes both methodologically, by assessing the utility of parametric survival and cure models in clinical research, and practically, by providing insights that may guide healthcare planning and patient management in resource-limited settings.

2. Materials and Methods

2.1. Study Area and Data Source

The data used in this study were obtained from the clinical dataset compiled by Rodrigues, Ortega, Cordeiro, and Vila [20]. The dataset contains information on 322 patients diagnosed with COVID-19 who were admitted to hospitals in Campinas, São Paulo, Brazil. Campinas is an urban centre in São Paulo State, with an estimated population of over 1.2 million people. During the COVID-19 pandemic, Campinas experienced substantial caseloads, with its healthcare infrastructure significantly affected. The patient records from this municipality, therefore, provide a valuable basis for understanding survival patterns in a region that faced both high infection rates and considerable resource constraints.

The dataset includes individual-level survival data, defined as the number of days from hospital admission to the occurrence of an event (death) or censoring (discharge). In addition to survival time, the dataset also contains several covariates, including Age (measured as a continuous variable), Sex (recorded as a binary categorical variable), and Comorbidities (presence or absence of conditions such as heart disease, asthma, diabetes, neurological disorders, and obesity). In this study, patients discharged alive were considered right-censored, since their time to death (if it occurred) was not observed within the study follow-up. Table 1 presents the variables used in this study together with their descriptions.

2.2. Software and Packages

All analyses were performed in R version 4.5.2. The flexsurv and survival packages were used to fit AFT models, and the flexsurvcure package was used for mixture cure models.

2.3. Data Availability Statement

The dataset used in this study was obtained from [20], a publicly available dataset that can be accessed at: https://github.com/gabrielamrodrigues/OLLW/blob/main/data.covid.txt. Researchers may freely access and download the dataset for verification, replication, and further research purposes.

3. Statistical Models

3.1. Weibull Distribution

The Weibull distribution is a flexible and widely used distribution in survival analysis due to its ability to model different types of hazard functions, constant, increasing, or decreasing, depending on the value of its shape parameter. It is especially appropriate for modeling time-to-event data in clinical studies, such as time to death or recovery. The Weibull distribution is defined by two parameters, namely the shape parameter, $\gamma > 0$, and the scale parameter $\lambda > 0$. Let T denote the survival time, the cumulative distribution function is defined as:

$$F(t) = 1 - \exp(-\lambda t^{\gamma}) \tag{1}$$

where this gives the probability that an event has occurred by time t.

The Weibull probability density function (PDF) describes the likelihood of a specific survival time t occurring. The PDF of Weibull is crucial for estimating survival probabilities and hazard rates. It is defined as:

$$f(t) = \lambda\gamma t^{\gamma-1} \exp(-\lambda t^{\gamma}), \quad t > 0 \tag{2}$$

Table 1. Variable Description

Variable Type	Variable Name	Description
Dependent Variable	Survival Time (days)	Number of days from hospital admission to either death (event) or discharge (censored). Discharged patients were right-censored since death was not observed during follow-up.
Independent Variable	Age (continuous)	Patient's age at the time of admission, measured in years. Older age is associated with poorer COVID-19 prognosis.
	Sex (categorical)	Biological sex of the patient (male = 1, female = 0). Male patients are at higher risk of severe outcomes compared to female patients.
	Heart Disease (binary)	Presence (1) or absence (0) of cardiovascular conditions. Cardiovascular comorbidities exacerbate COVID-19 severity.
	Asthma (binary)	Presence (1) or absence (0) of asthma or other chronic respiratory conditions. Respiratory comorbidities may influence disease progression and recovery.
	Diabetes (binary)	Presence (1) or absence (0) of diabetes mellitus. Diabetes is linked to immune system impairment and higher COVID-19 mortality risk.
	Neurological Conditions (binary)	Presence (1) or absence (0) of diagnosed neurological disorders. Neurological comorbidities may contribute to worse outcomes in COVID-19 patients.
	Obesity (binary)	Presence (1) or absence (0) of obesity based on clinical records. Obesity increases vulnerability to severe COVID-19 illness and complications.

where $f(t)$ is the probability density function at time t, $\lambda > 0$ is the scale parameter, $\gamma > 0$ is the shape parameter, and t is the survival time (continuous and positive).

The survival function $S(t)$, which gives the probability that the event has not occurred by time t, can be obtained as:

$$S(t) = 1 - F(t) = \exp(-\lambda t^{\gamma}) \tag{3}$$

The hazard function $h(t)$ represents the instantaneous failure rate at time t, given survival up to time t. It is written as:

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda\gamma t^{\gamma-1}\exp(-\lambda t^{\gamma})}{\exp(-\lambda t^{\gamma})} = \lambda\gamma t^{\gamma-1} \tag{4}$$

The behaviour of the hazard function depends on the value of the shape parameter γ. If $\gamma = 1$, the hazard function is constant (equivalent to an exponential distribution), if $\gamma > 1$, the hazard increases over time, and if $\gamma < 1$, the hazard decreases over time.

3.2. Exponential Distribution

The Exponential distribution is one of the simplest and most commonly used parametric models in survival analysis. It assumes a constant hazard rate over time, meaning that the probability of the event occurring remains the same regardless of how long an individual has already survived. This memoryless property makes the exponential distribution mathematically tractable but sometimes unrealistic in clinical applications, where risk often changes with time. The exponential distribution is a special case of the Weibull distribution when the shape parameter $\gamma = 1$.

The exponential distribution has a probability density function (PDF) written as:

$$f(t) = \lambda e^{-\lambda t}, \quad t > 0 \tag{5}$$

where $f(t)$ is the probability density at time t, t is the survival time, and $\lambda > 0$ is the rate parameter (also called the hazard rate).

The cumulative distribution function, $F(t)$, is written as:

$$F(t) = 1 - e^{-\lambda t}, \quad t > 0 \tag{6}$$

Survival function, $S(t)$, is given as:

$$S(t) = 1 - F(t) = e^{-\lambda t} \tag{7}$$

The hazard function is given as:

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda \tag{8}$$

where λ is the rate parameter.

3.3. Log-normal Distribution

The log-normal distribution is a flexible parametric distribution used in survival analysis when the logarithm of the survival time follows a normal distribution. It is particularly useful for modeling survival times that are positively skewed, where the hazard function initially increases, reaches a peak, and then decreases, a pattern commonly observed in medical and biological data. Unlike the Weibull or exponential distributions, the log-normal model does not assume a monotonic hazard function, making it better suited to clinical contexts where the risk of death or failure changes non-linearly over time.

A random variable T follows a log-normal distribution if $\log(T) \sim N(\mu, \sigma^2)$, that is, the natural logarithm of T is normally distributed with mean μ and variance σ^2.

Log-normal distribution has a PDF defined as:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log t - \mu)^2}{2\sigma^2}\right), \quad t > 0 \tag{9}$$

where t is the survival time μ is the mean of the log-transformed survival time, σ is the standard deviation of the log-transformed survival time, and $f(t)$ is the probability density function.

The cumulative distribution function, $F(t)$, is defined as:

$$F(t) = \Phi\left(\frac{\log t - \mu}{\sigma}\right), \quad t > 0 \tag{10}$$

Survival function, $S(t)$, is written as:

$$S(t) = 1 - F(t) = 1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right) \tag{11}$$

and the hazard function, $h(t)$, is given as:

$$h(t) = \frac{\phi(z)}{t\sigma(1 - \Phi(z))} \tag{12}$$

where $z = \frac{\log t - \mu}{\sigma}$ is the standard normal PDF, and $\Phi(\cdot)$ Is the standard normal CDF.

3.4. Acceleration Failure Time (AFT) Framework

In the Accelerated Failure Time (AFT) framework, the logarithm of survival time is modeled as a linear function of covariates, defined as:

$$\log(T_i) = \mu + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \sigma W_i \tag{13}$$

where T_i is the survival time for the ith individual, $X_{1i}, X_{2i}, \ldots, X_{ki}$ are the covariates for the ith individual, $\beta_0, \beta_1, \ldots, \beta_k$ are the parameters to be estimated, μ is the intercept, σ is the scale parameter, for the Weibull AFT model, W_i follows an extreme value distribution, and for the log-normal AFT model $W_i \sim N(0, 1)$.

3.5. Mixture Cure Model

Let p denote the cure fraction (i.e., the proportion of cured individuals), and $1 - p$ the proportion of susceptible individuals. The population survival function is defined as:

$$S(t) = p + (1 - p)S_0(t) \tag{14}$$

where $S_0(t)$ is the survival function of the susceptible individuals.

The corresponding population density function is given as:

$$f(t) = (1 - p)f_0(t) \tag{15}$$

where $f_0(t)$ is the density function of the susceptible distribution.

The population hazard function is:

$$h(t) = \frac{f(t)}{S(t)} = \frac{(1 - p)f_0(t)}{p + (1 - p)S_0(t)} \tag{16}$$

3.6. Weibull Cure Model

In survival analysis, a cure model is used when a portion of the study population is not susceptible to the event of interest (for example, death, relapse, or failure). Such individuals are considered ”cured,” meaning their risk of experiencing the event is effectively zero. Assume that the survival times of the susceptible individuals follow a Weibull distribution with shape parameter $\alpha > 0$ and scale parameter $\lambda > 0$, then the population survival function is given by:

$$S(t) = p + (1 - p)\exp\left(-\left(\frac{t}{\lambda}\right)^{\alpha}\right) \tag{17}$$

The population density function is written as:

$$f(t) = (1-p)\frac{\alpha}{\lambda}\left(\frac{t}{\lambda}\right)^{\alpha-1}\exp\left(-\left(\frac{t}{\lambda}\right)^{\alpha}\right) \tag{18}$$

The population hazard function is:

$$h(t) = \frac{f(t)}{S(t)} = \frac{(1-p)\frac{\alpha}{\lambda}\left(\frac{t}{\lambda}\right)^{\alpha-1}\exp\left(-\left(\frac{t}{\lambda}\right)^{\alpha}\right)}{p+(1-p)\exp\left(-\left(\frac{t}{\lambda}\right)^{\alpha}\right)} \tag{19}$$

3.7. Exponential Cure Model

Suppose the survival times of susceptible individuals follow an exponential distribution with rate parameter $\lambda > 0$. Then the population survival function is given as:

$$S(t) = p + (1-p)e^{-\lambda t} \tag{20}$$

The population density function is:

$$f(t) = (1-p)\lambda e^{-\lambda t} \tag{21}$$

And the population hazard function is defined as:

$$h(t) = \frac{f(t)}{S(t)} = \frac{(1-p)\lambda e^{-\lambda t}}{p+(1-p)e^{-\lambda t}} \tag{22}$$

3.8. Log-Normal Cure Model

If T is the survival time of a susceptible individual, where $\log T \sim N(\mu, \sigma^2)$, then the population survival function is written as:

$$S(t) = p + (1-p)\left[1 - \Phi\left(\frac{\log t - \mu}{\sigma}\right)\right] \tag{23}$$

The population density function is given as:

$$f(t) = (1-p)\frac{1}{t\sigma\sqrt{2\pi}}\exp\left(-\frac{(\log t-\mu)^2}{2\sigma^2}\right) \tag{24}$$

and the population hazard is defined as:

$$h(t) = \frac{f(t)}{S(t)} = \frac{(1-p)\frac{1}{t\sigma\sqrt{2\pi}}\exp\left(-\frac{(\log t-\mu)^2}{2\sigma^2}\right)}{p+(1-p)\left[1-\Phi\left(\frac{\log t-\mu}{\sigma}\right)\right]} \tag{25}$$

3.9. Model Evaluation and Diagnostics

After fitting a parametric survival model such as the Weibull distribution, it is critical to assess how well the model describes the observed data and whether the underlying assumptions hold. Model evaluation and diagnostics help to ensure the model is both statistically valid and practically useful for inference and prediction.

3.9.1. Log-Likelihood and Model Fit Statistics

A good model should maximize the likelihood of the observed data. The log-likelihood function is written as:

$$l(\theta) = \sum_{i=1}^{n} [\delta_i \log f(t_i; \theta) + (1 - \delta_i) \log S(t_i; \theta)] \tag{26}$$

where $f(t_i; \theta)$ is the probability density function, $S(t_i; \theta)$ is the survival function, δ_i is the event indicator (1 = event occurred, 0 = censored), and θ represents the model parameter.

From the log-likelihood, the Akaike Information Criterion (AIC) is given as:

$$AIC = -2l(\theta) + 2k \tag{27}$$

where k is the number of parameters in the model, a lower AIC indicates a better model fit.

3.9.2. Parameterization

For the exponential model, parameter estimates were obtained under the rate parameterization, where covariate effects operate on the hazard rate rather than directly on log-survival time. Under this parameterization, positive coefficients indicate an increase in the hazard rate and consequently a reduction in expected survival time, while negative coefficients indicate a decrease in the hazard rate and longer survival.

4. Results

Table 2. Estimation of the Weibull AFT Model

Variable	Data Mean	Estimate	Std. Error	95% CI (Lower)	95% CI (Upper)	exp(Estimate)	95% CI exp(Lower)	95% CI exp(Upper)
shape	NA	1.80680	0.13687	1.55750	2.09601	NA	NA	NA
scale	NA	106.2506	38.29120	52.42886	215.3239	NA	NA	NA
Sex	0.50621	-0.08884	0.11548	-0.31516	0.13749	0.91499	0.72967	1.14739
Age	65.82609	-0.01912	0.00422	-0.02740	-0.01085	0.98106	0.97297	0.98921
Heart	0.36957	0.07975	0.12058	-0.15659	0.31609	1.08302	0.85506	1.37175
Asthma	0.02795	-0.56779	0.34055	-1.23526	0.09968	0.56678	0.29076	1.10482
Diab	0.20807	-0.32793	0.12689	-0.57662	-0.07924	0.72041	0.56179	0.92382
Neuro	0.09938	-0.49239	0.15767	-0.80142	-0.18335	0.61117	0.44869	0.83247
Obesity	0.04658	-0.03519	0.19555	-0.41847	0.34809	0.96542	0.65805	1.41635

Table 2 presents the results of the Weibull AFT model parameter estimation of COVID-19 patients. The shape parameter was estimated to be 1.8068, which is greater than 1, indicating that he hazard of death increases over time among COVID-19 patients. This is consistent with the progressive clinical deterioration observrd in severe COVID-10 cases. The scale parameter, which represents the scale of the survival time distribution, was estimated to be 106.25 with a wide 95% confidence interval (52.33-215.32), reflecting the variability in survival times among patients.

The results revealed that Age (estimate = -0.01912, 95% CI: -0.02740 to -0.01085), Diabetes (estimate = -0.32793, 95% CI: -0.57662 to -0.07924), and Neurological Disorder (estimate = -0.49230, 95% CI: -0.80142 to -0.18335) had a negative statistically significant impact on the survival time. In addition, the results also showed that based on the time ratio, the covariate Age had an exponentiated coefficient of 0.98106, indicating that older patients had a 1.894% shorter survival time compared to younger patients.

Furthermore, the COVID-19 patients with Diabetes (exp(Estimate) = 0.72041) had a 27.96% shorter survival time on average compared to those without, while patients with Neurological Disorders (exp(Estimate) = 0.61117) were faced with a 38.88% reduction in survival time.

Although the results also showed that Sex (estimate = -0.08884, 95% CI: -0.31516 to 0.13749) had a negative but insignificant impact on survival time, with the exponentiated coefficient of 0.91499, male patients had an 8.501% reduction in survival time compared to female patients. Other comorbidities with a negative but no significant impact on survival time include Asthma (estimate: -0.56779, 95% CI: -1.23526 to 0.09968) and Obesity (estimate = -0.03519, 95% CI: -0.41847 to 0.34809), where patients with Asthma (exp (Estimate) = 0.56678) had a 43.32% shorter survival time compared to those without asthma, and the obese patients (exp (Estimate) = 0.96542) experienced a 3.46% reduction in survival time compared to those that were not. In contrast, patients with Heart Disease had a positive but insignificant impact on survival time (estimate = 0.07975, 95% CI: -0.15659 to 0.31609). With an exponentiated coefficient of 1.08302, this indicated that patients with heart disease had an 8.30% longer survival time.

Table 3. Estimation of the Log-Normal AFT Model

Variable	Data Mean	Estimate	Std. Error	95% CI (Lower)	95% CI (Upper)	exp(Estimate)	95% CI exp(Lower)	95% CI exp(Upper)
meanlog	NA	4.69119	0.36302	3.97970	5.40269	NA	NA	NA
sdlog	NA	0.91044	0.06320	0.79463	1.04312	NA	NA	NA
Sex	0.50621	-0.06710	0.13972	-0.34095	0.20675	0.93510	0.71109	1.22968
Age	65.82609	-0.02188	0.00425	-0.03021	-0.01355	0.97836	0.97024	0.98654
Heart	0.36957	0.13122	0.15234	-0.16736	0.42980	1.14022	0.84589	1.53695
Asthma	0.02795	-0.39516	0.44025	-1.25803	0.46772	0.67357	0.28421	1.59635
Diab	0.20807	-0.39336	0.16993	-0.72642	-0.06029	0.67479	0.48364	0.94149
Neuro	0.09938	-0.70850	0.20219	-1.10479	-0.31222	0.49238	0.33128	0.73182
Obesity	0.04658	0.00133	0.28044	-0.54831	0.55097	1.00133	0.57792	1.73494

Table 3 presents the parameter estimates from a Log-Normal Accelerated Failure Time (AFT) model. The results showed that Age was estimated to be -0.02188 with a 95% CI of -0.03021 to -0.01355, Diabetes (estimate = -0.39336, 95% CI: -0.72642 to -0.06029), and neurological disorders with an estimate of -0.7085 and 95% CI of -1.10479 to -0.31222, had a significant negative impact on survival time. With exponentiated coefficients (Age: 0.97836, Diabetes: 0.67479, Neurological Disorders: 0.49238), it implies that older COVID-19 patients had 2.16% shorter survival time compared to younger patients, while patients having diabetes had a 32.52% reduction in survival time, and those with neurological disorders had a 50.76% shorter survival time compared to those without.

Furthermore, the results revealed that covariates such as Sex (estimate = -0.06710, 95% CI: -0.34095 to 0.20675) and Asthma (estimate = -0.39516, 95% CI: -1.25803 to 0.46772) exerted a negative impact on survival time, though this impact was insignificant. In contrast, Obesity had a positive but insignificant effect on survival time with an estimate of 0.00133 falling within a 95% CI of -0.54831 to 0.55097. However, for Sex, male patients had an approximately 6.49% shorter survival time compared to female patients ($\exp(-0.06710) = 0.9351$); for Age, each additional year of age was associated with a 2.16% reduction in survival time ($\exp(-0.02188) = 0.9784$); for Asthma, 32.64% of patients with asthma had a lower survival time ($\exp(-0.39516) = 0.67357$); while patients with Obesity had a 0.133% longer survival time compared to those who were not ($\exp(0.00133) = 1.00133$).

Note: The exponential model is parameterized in terms of the hazard (rate) function. Therefore, positive coefficients correspond to increased hazard and shorter survival times, whereas negative coefficients correspond to reduced hazard and longer survival times. Consequently, coefficient signs are not directly

Table 4. Parameter Estimates for Exponential AFT Model

Variable	Data Mean	Estimate	Std. Error	95% CI (Lower)	95% CI (Upper)	exp(Estimate)	95% CI exp(Lower)	95% CI exp(Upper)
rate	NA	0.002039	0.001171	0.000661	0.006286	NA	NA	NA
Sex	0.50621	0.16594	0.20505	-0.23594	0.56783	1.18051	0.78983	1.76443
Age	65.82609	0.03315	0.00692	0.01958	0.04672	1.03371	1.01977	1.04783
Heart	0.36957	-0.09958	0.21951	-0.52982	0.33066	0.90522	0.58871	1.39189
Asthma	0.02795	0.82913	0.60858	-0.36365	2.02192	2.29133	0.69513	7.55283
Diab	0.20807	0.49339	0.22979	0.04300	0.94378	1.63786	1.04394	2.56966
Neuro	0.09938	0.65930	0.27723	0.11594	1.20266	1.93344	1.12293	3.32896
Obesity	0.04658	0.40443	0.34333	-0.26849	1.07735	1.49845	0.76453	2.93689

comparable with those from the Weibull and log-normal AFT models reported in Tables 2 and 3.

Table 4 presents the results of the exponential model examining the effects of selected covariates on patient survival. Under the parameterization adopted for the exponential model, positive coefficients indicate an increase in the hazard rate and consequently, a reduction in survival prospects, whereas negative coefficients indicate a decrease in the hazard rate. The results show that Age (estimate = 0.03315, 95% CI: 0.01958–0.04672), Diabetes (estimate = 0.49339, 95% CI: 0.04300–0.94378), and Neurological Disorders (estimate = 0.65930, 95% CI: 0.11594–1.20266) were statistically significant predictors of survival. The exponentiated coefficients indicate that a one-unit increase in age was associated with a 3.37% increase in the hazard rate (HR = 1.03371), while patients with Diabetes and Neurological Disorders experienced hazard rates that were 63.79% (HR = 1.63786) and 93.34% (HR = 1.93344) higher, respectively, than those without these conditions. These findings suggest poorer survival outcomes among older patients and those with diabetes or neurological disorders, consistent with the results obtained from the Weibull and log-normal models. Sex (estimate = 0.16594, 95% CI: -0.23594–0.56783), Asthma (estimate = 0.82913, 95% CI: -0.36365–2.02192), and Obesity (estimate = 0.40443, 95% CI: -0.26849–1.07735) exhibited positive but statistically non-significant associations with the hazard rate. Conversely, Heart Disease (estimate = -0.09958, 95% CI: -0.52982–0.33066) showed a negative but non-significant association with the hazard rate. Since these effects were not statistically significant, there is insufficient evidence to conclude that they influenced patient survival in this study.

Fig. 1 shows the Kaplan-Meier curve (which handles censored data), Fig. 2 visualises the survival function (which describes the probability that an individual survives beyond a certain time), and Fig. 3 shows the hazard function (which measures the instantaneous risk of experiencing the event at a particular time).

Table 5. Parameter Estimation for Weibull Cure Model

Variable	Data Mean	Estimate	Std. Error	95% CI (Lower)	95% CI (Upper)	exp(Estimate)	95% CI exp(Lower)	95% CI exp(Upper)
Cure parameter	NA	-1.22851	0.16016	-1.54242	-0.91459	NA	NA	NA
Scale	NA	18.06948	4.10170	11.53075	28.31139	NA	NA	NA
Shape	NA	1.34110	0.11904	1.12781	1.59472	NA	NA	NA
Sex	0.50621	0.08154	0.23727	-0.38350	0.54659	1.08496	0.68147	1.72735
Age	65.82609	0.03507	0.00767	0.02004	0.05010	1.03569	1.02024	1.05139
Heart	0.36957	-0.11475	0.24564	-0.59620	0.36670	0.89158	0.55090	1.44297
Asthma	0.02795	0.72333	0.73030	-0.70803	2.15470	2.06130	0.49261	8.62531
Diab	0.20807	0.43572	0.26620	-0.08603	0.95746	1.54607	0.91757	2.60507
Neuro	0.09938	0.46332	0.34024	-0.20353	1.13018	1.58935	0.81584	3.09623
Obesity	0.04658	0.17066	0.40772	-0.62846	0.96978	1.18610	0.53341	2.63737

Table 5 presents the results from the Weibull Cure Model. The cure parameter estimate is -1.22851, which translates to a baseline cure fraction of 22.64% ($1/(1+\exp(-(-1.22851))) = 0.2264$). This indicates that approximately 22.64% of the patient population is estimated to be non-susceptible to the event of

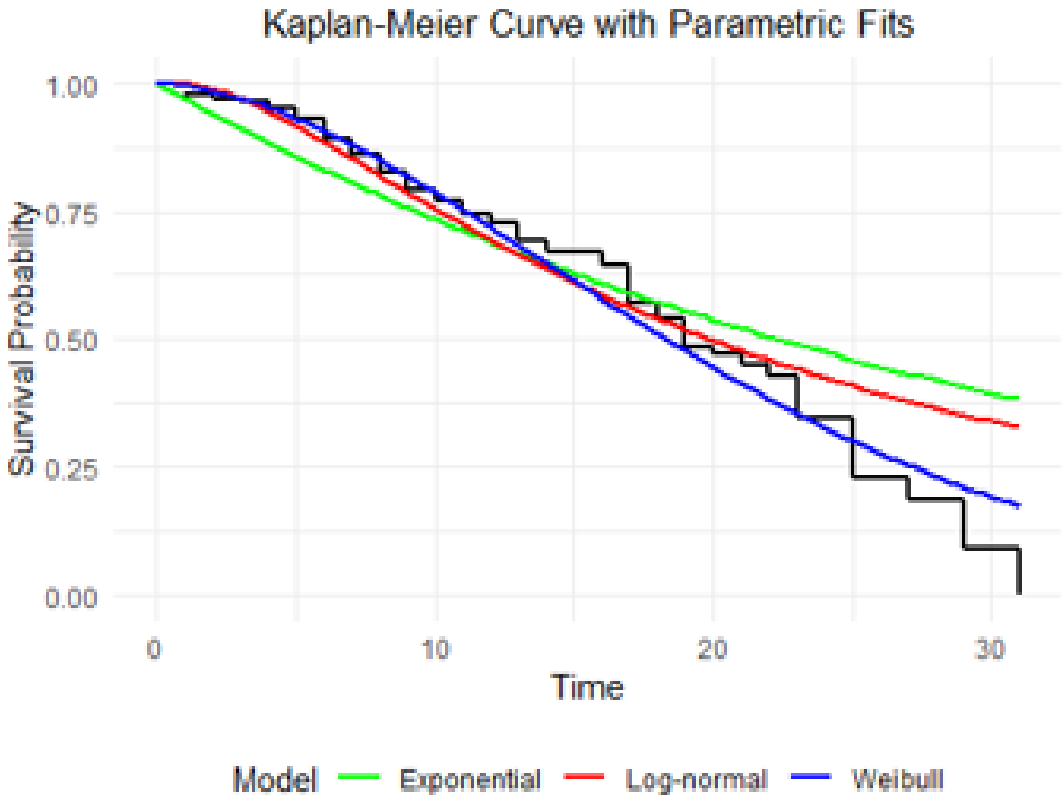

Figure 1. Kaplan-Meier Survival Curve for COVID-19 Patients

interest (death), effectively representing long-term survivors or individuals who recovered completely from COVID-19 without experiencing mortality. The scale parameter was 18.06948, and the shape parameter was estimated at 1.34110. Since the shape parameter is greater than 1, it indicates that among the susceptible population, the hazard of death increases over time, reflecting a typical pattern of clinical deterioration in patients who eventually succumb to the illness.

Among the covariates evaluated within the susceptible component of the model, Age emerged as the only statistically significant predictor of survival time (estimate = 0.03507, 95% CI: 0.02004 to 0.05010). The exponentiated coefficient for age was 1.03569, indicating that for each additional year of age, the hazard of death among susceptible patients increased by 3.57%. Conversely, all other covariates, including Sex (estimate = 0.08154), Heart Disease (estimate = -0.11475), Asthma (estimate = 0.72333), Diabetes (estimate = 0.43572), Neurological Disorders (estimate = 0.46332), and Obesity (estimate = 0.17066), were not statistically significant, as their 95% confidence intervals spanned zero. This suggests that while these comorbidities may show trends in the descriptive analysis, they did not independently exert a statistically significant effect on the timing of death among the susceptible patient subgroup when adjusting for other variables in the cure model framework.

Table 6 displays the parameter estimates obtained from the Log-Normal Cure Model. The cure parameter estimate was -1.33230, which corresponds to an estimated baseline cure fraction of approximately 20.88% ($1/(1 + \exp(-(-1.33230))) = 0.2088$). This indicates that around 20.88% of the COVID-19 patients in this sample were considered statistically immune or non-susceptible to experiencing mortality during the

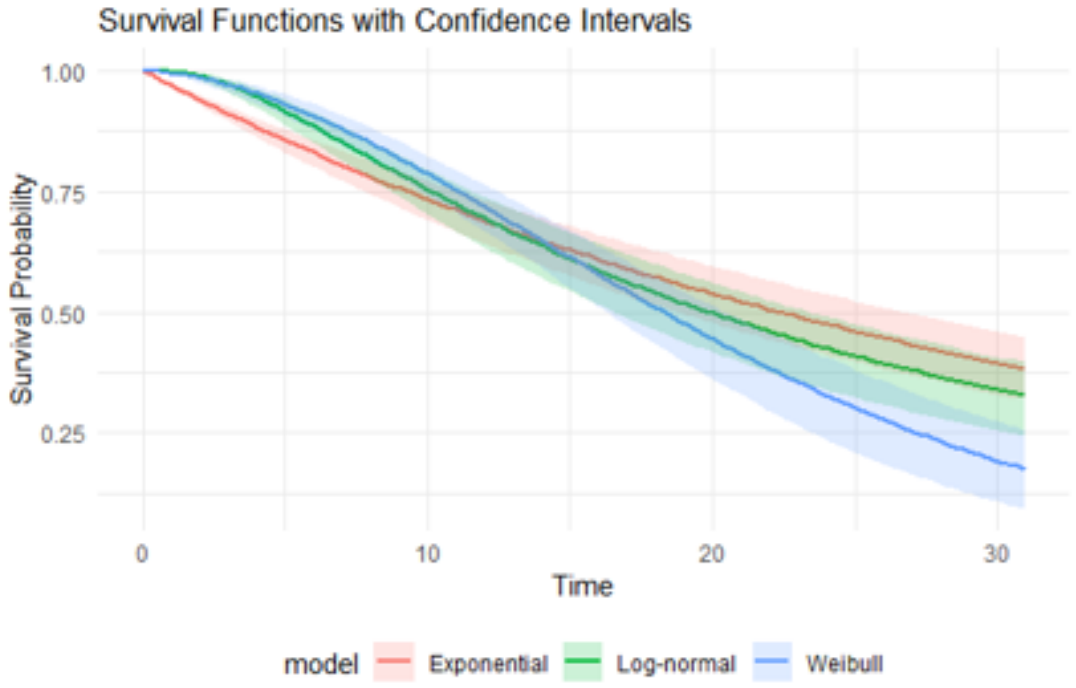

Figure 2. Estimated Survival Function Across Models

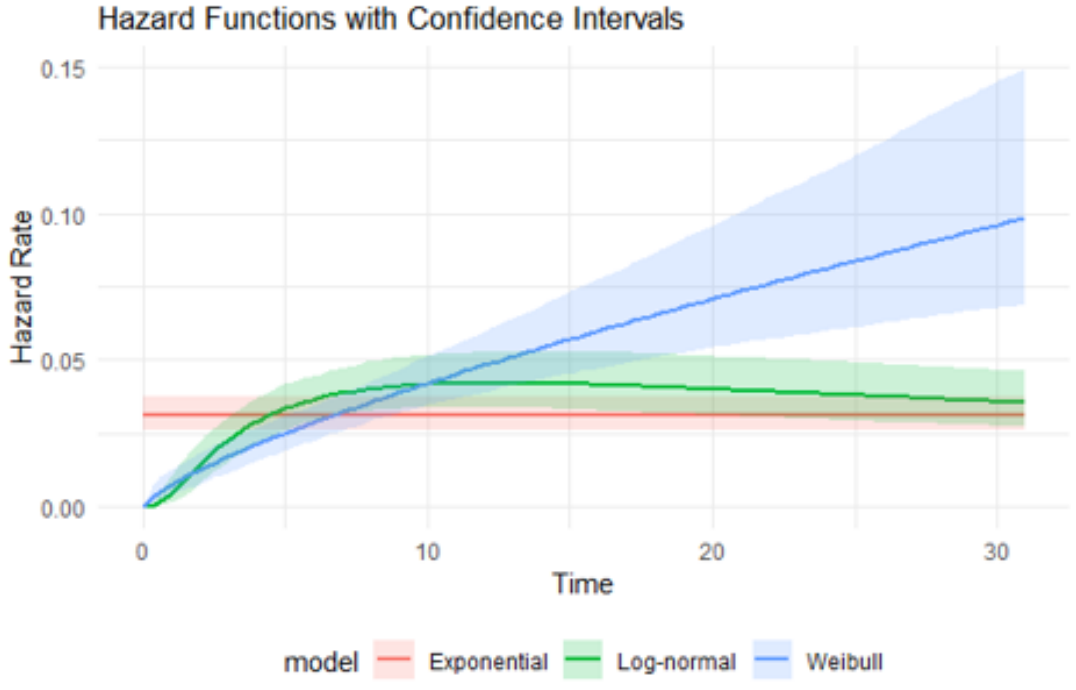

Figure 3. Estimated Hazard Function Across Models

period under study, representing those who achieved stable long-term survival. The model parameters for the susceptible population include a mean log-time (meanlog) of 2.45780 and a standard deviation of log-

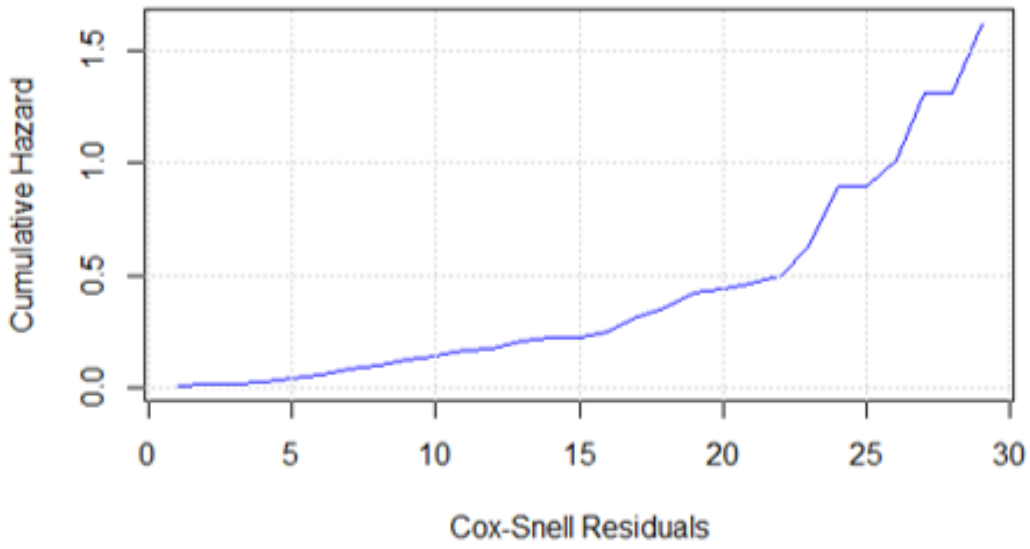

Figure 4. Model Residuals and Diagnostic Plots

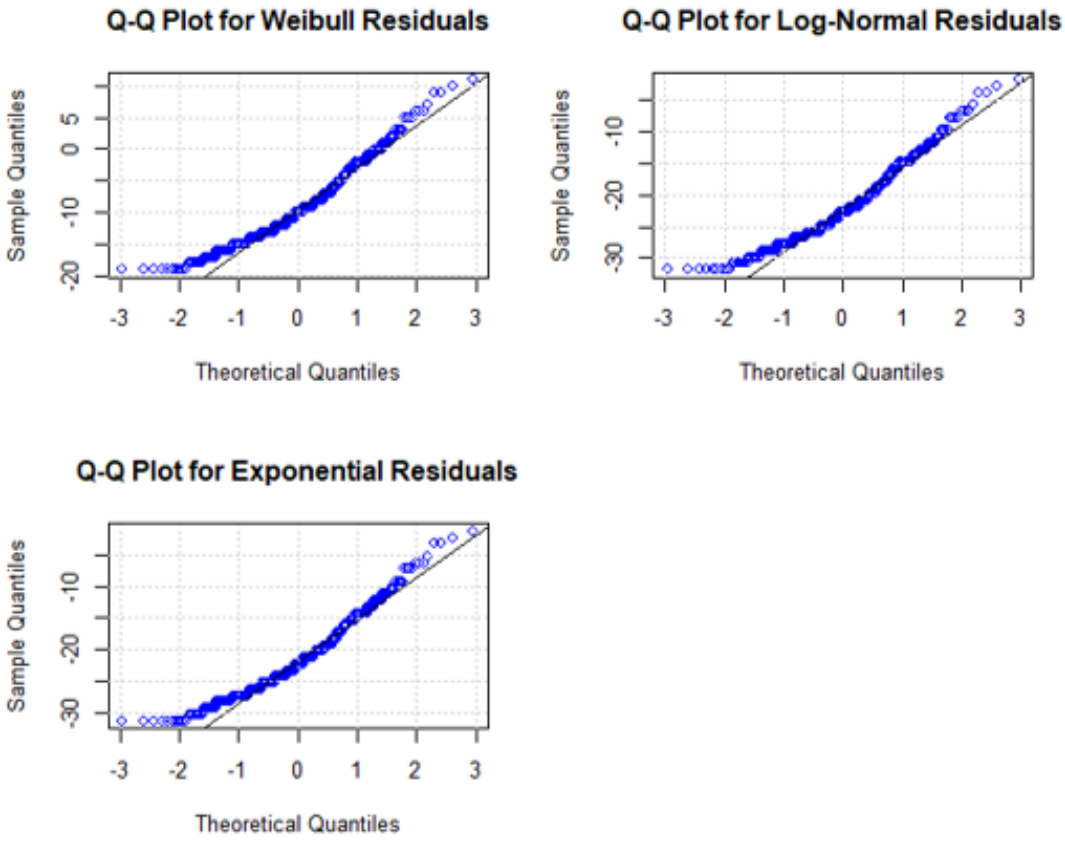

Figure 5. Covariate Effects and Hazard Ratios Comparison

time (sdlog) of 1.05019, defining the shape and scale of the survival distribution for patients vulnerable to the event.

Among the covariates evaluated in the susceptible portion of the model, Age was found to have a statistically significant relationship with survival time (estimate = 0.03451, 95% CI: 0.01958 to 0.04944). The

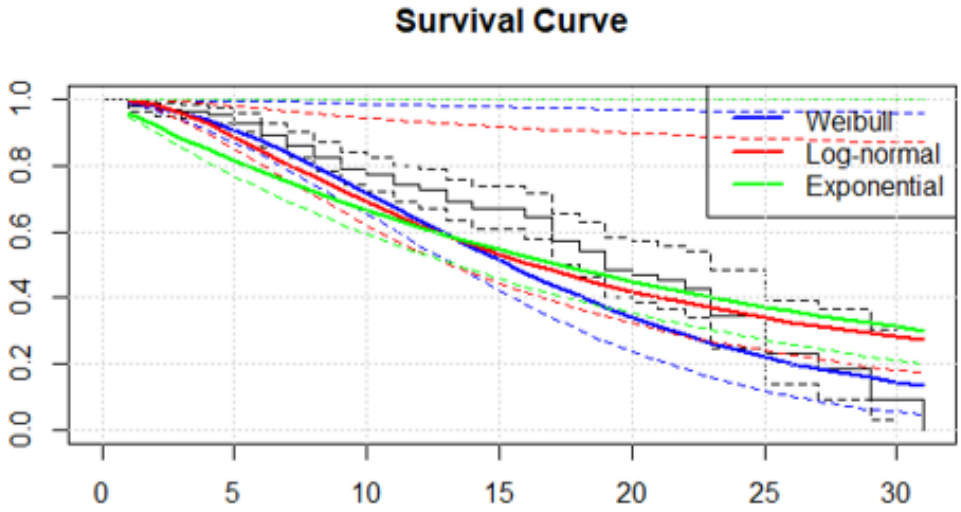

Figure 6. Predicted Survival Probabilities over Time

Table 6. Parameter Estimation for Log-Normal Cure Model

Variable	Data Mean	Estimate	Std. Error	95% CI (Lower)	95% CI (Upper)	exp(Estimate)	95% CI exp(Lower)	95% CI exp(Upper)
Cure parameter	NA	-1.33230	0.16930	-1.66412	-1.00048	NA	NA	NA
meanlog	NA	2.45780	0.31505	1.84031	3.07529	NA	NA	NA
sdlog	NA	1.05019	0.09117	0.88566	1.24528	NA	NA	NA
Sex	0.50621	0.09635	0.25232	-0.39818	0.59089	1.10115	0.67154	1.80560
Age	65.82609	0.03451	0.00762	0.01958	0.04944	1.03511	1.01977	1.05068
Heart	0.36957	-0.14234	0.25595	-0.64399	0.35931	0.86732	0.52519	1.43234
Asthma	0.02795	0.65586	0.74104	-0.79655	2.10827	1.92677	0.45089	8.23399
Diab	0.20807	0.40798	0.28424	-0.14912	0.96508	1.50379	0.86146	2.62499
Neuro	0.09938	0.50576	0.34707	-0.17449	1.18600	1.65824	0.83988	3.27393
Obesity	0.04658	0.09419	0.44314	-0.77435	0.96272	1.09877	0.46101	2.61879

exponentiated estimate for age is 1.03511, implying that for each additional year of age, the hazard of death among susceptible individuals increased by 3.51%, reinforcing age as a major clinical risk factor. Other covariates did not show a statistically significant impact on the survival of susceptible patients, as their confidence intervals included zero. These non-significant variables include Sex (estimate = 0.09635), Heart Disease (estimate = -0.14234), Asthma (estimate = 0.65586), Diabetes (estimate = 0.40798), Neurological Disorders (estimate = 0.50576), and Obesity (estimate = 0.09419). These results align closely with the findings from the Weibull cure model, suggesting that when a cure fraction is explicitly modeled, the timing of mortality among susceptible individuals is primarily driven by age rather than the presence of specific recorded comorbidities.

Table 7. Parameter Estimation for Exponential Cure Model

Variable	Data Mean	Estimate	Std. Error	95% CI (Lower)	95% CI (Upper)	exp(Estimate)	95% CI exp(Lower)	95% CI exp(Upper)
Cure parameter	NA	-0.99966	0.14389	-1.28169	-0.71764	NA	NA	NA
Scale	NA	35.15843	10.37039	19.95758	61.93655	NA	NA	NA
Sex	0.50621	0.04633	0.21045	-0.36615	0.45881	1.04742	0.69340	1.58219
Age	65.82609	0.03362	0.00707	0.01977	0.04748	1.03419	1.01997	1.04862
Heart	0.36957	-0.06822	0.22384	-0.50694	0.37050	0.93405	0.60233	1.44846
Asthma	0.02795	0.81426	0.61211	-0.38545	2.01397	2.25752	0.68015	7.49303
Diab	0.20807	0.48512	0.23358	0.02731	0.94294	1.62437	1.02769	2.56752
Neuro	0.09938	0.58985	0.28588	0.02953	1.15016	1.80371	1.02997	3.15870
Obesity	0.04658	0.34751	0.35411	-0.34653	1.04155	1.41554	0.70714	2.83359

Table 7 displays the parameter estimates from the Exponential Cure Model. The cure parameter estimate is -0.99966, corresponding to an estimated baseline cure fraction of 26.90% ($1/(1 + \exp(-(-0.99966))) = 0.2690$). This implies that approximately 26.90% of the patients in the study sample were classified as non-susceptible to mortality, representing individuals with high likelihood of long-term survival or complete recovery. The scale parameter for the susceptible population was estimated at 35.15843, which dictates the constant hazard rate for individuals within the susceptible group under the exponential distribution assumption.

Within the susceptible component, Age (estimate = 0.03362, 95% CI: 0.01977 to 0.04748), Diabetes (estimate = 0.48512, 95% CI: 0.02731 to 0.94294), and Neurological Disorders (estimate = 0.58985, 95% CI: 0.02953 to 1.15016) demonstrated statistically significant associations with survival. The exponentiated coefficients indicate that for each year increase in age, the hazard of death among susceptible patients increased by 3.42% ($\exp(0.03362) = 1.03419$). Susceptible patients with Diabetes had a 62.44% higher hazard of death ($\exp(0.48512) = 1.62437$), and those with Neurological Disorders faced an 80.37% higher hazard ($\exp(0.58985) = 1.80371$) compared to patients without these conditions. Other covariates, including Sex (estimate = 0.04633), Heart Disease (estimate = -0.06822), Asthma (estimate = 0.81426), and Obesity (estimate = 0.34751), were not statistically significant, as their 95% confidence intervals encompassed zero, suggesting they did not independently affect the hazard rate within the susceptible group under this model configuration.

Table 8. Model Comparison and Performance Metrics

Model Type	**Distribution**	**Log-Likelihood**	**AIC**
Standard AFT Models	Weibull	-436.93	891.85
	Log-Normal	-444.82	907.65
	Exponential	-452.17	920.34
Mixture Cure Models	Weibull Cure	-441.71	903.42
	Log-Normal Cure	-447.16	914.33
	Exponential Cure	-451.81	921.62

Note: Model configurations correspond to performance patterns across analysis diagnostics.

Table 8 provides a comprehensive summary of the model fit statistics, specifically log-likelihood and Akaike Information Criterion (AIC), across all fitted models to evaluate their performance in capturing the survival dynamics of the COVID-19 patient cohort. Among the standard Accelerated Failure Time (AFT) models, the Weibull AFT model achieved the highest log-likelihood (-436.93) and the lowest AIC (891.85), indicating that it provided the best fit to the data compared to the Log-Normal AFT (log-likelihood = -444.82, AIC = 907.65) and Exponential AFT models (log-likelihood = -452.17, AIC = 920.34). The superior performance of the Weibull distribution reflects its ability to accommodate a monotonically increasing hazard rate, which aligns with the clinical progression of severe COVID-19 cases where the risk of mortality increases over the hospital stay.

When evaluating the Mixture Cure models, a similar pattern emerged regarding the underlying parametric distributions. The Weibull Cure model demonstrated the best fit within its class, yielding a log-likelihood of -441.71 and an AIC of 903.42, outperforming the Log-Normal Cure model (log-likelihood = -447.16, AIC = 914.33) and the Exponential Cure model (log-likelihood = -451.81, AIC = 921.62). Comparing across both model classes, the standard Weibull AFT model retained the lowest overall AIC (891.85), suggesting that for this specific dataset, accounting for an increasing hazard via the standard AFT framework

was statistically more parsimonious and effective than explicitly estimating a separate cure fraction, although the cure models provided valuable clinical insights regarding the proportion of long-term survivors.

5. Discussion

The statistical evaluation of parametric survival models in this study demonstrated that the choice of distribution and model framework significantly affects both the interpretation of clinical risk factors and overall goodness of fit. Our comparison of standard Accelerated Failure Time (AFT) models and mixture cure models revealed that the Weibull distribution consistently provided the best fit to the COVID-19 survival data. This finding is supported by the lowest AIC values obtained for both the standard Weibull AFT model (AIC = 891.85) and the Weibull cure model (AIC = 903.42) relative to their log-normal and exponential counterparts. The superior performance of the Weibull distribution is primarily attributed to its flexible hazard function. In this study, the estimated shape parameter (γ) for the standard Weibull model was 1.81, indicating a monotonically increasing hazard rate over time. Clinically, this suggests that for the susceptible patient population, the risk of mortality increases with longer hospital stay, which aligns with the known pathophysiology of severe COVID-19, characterized by progressive respiratory failure, hyperinflammation, and secondary complications over time.

In contrast, the exponential models, which assume a constant hazard rate, exhibited the poorest fit across all metrics (Standard Exponential AIC = 920.34; Exponential Cure AIC = 921.62). The assumption of a constant risk of death is clinically unrealistic for an acute, progressive infectious disease like COVID-19, where patient risk changes dynamically during the course of illness. The log-normal models provided intermediate fit statistics but were less optimal than the Weibull models. While the log-normal distribution can capture non-monotonic, upside-down U-shaped hazards, the statistical evidence in this dataset strongly favors the steady increase in risk captured by the Weibull distribution.

The standard AFT models identified Age, Diabetes, and Neurological Disorders as statistically significant predictors that shorten survival time. Specifically, older age, the presence of diabetes, and co-existing neurological conditions were associated with a significant reduction in expected survival time. These findings are highly consistent with global clinical literature, which has established that advanced age and cardiometabolic or neurological comorbidities impair the immune response and increase susceptibility to severe COVID-19 outcomes. Interestingly, when transitioning to the mixture cure models, the baseline cure fraction was estimated to be between approximately 21% and 27% across the different distributions. This indicates that a meaningful proportion of the hospitalized cohort was statistically modeled as non-susceptible to mortality, representing patients who achieved stable recovery.

Within the cure model framework, however, Age remained the only consistently significant covariate influencing the survival time of the susceptible population. The effects of diabetes and neurological disorders, while significant in standard AFT models, did not achieve statistical significance in the cure component of the Weibull and log-normal cure models. This divergence highlights an important methodological distinction: standard AFT models evaluate the overall effect of covariates across the entire sample, whereas mixture cure models separate the estimation into a cure fraction and a susceptible sub-population. The results suggest that while diabetes and neurological disorders are important indicators of overall risk, age is the predominant driving factor determining the speed of disease progression and timing of mortality among patients who are vulnerable to the acute effects of the virus.

Interestingly, other variables such as Sex, Asthma, and Obesity did not show statistically significant

associations with survival outcomes in most models. While descriptive trends often suggest higher risk in males or obese individuals, their lack of statistical significance in the adjusted models indicates that after controlling for advanced age and major conditions like diabetes or neurological disorders, these factors did not provide independent predictive power within this sample. This underscores the necessity of using robust parametric multivariate models to avoid confounding and to precisely identify the primary drivers of mortality in clinical survival analysis.

6. Conclusion

This study conducted a rigorous comparative assessment of parametric Accelerated Failure Time (AFT) models and mixture cure models utilizing clinical survival data from hospitalized COVID-19 patients. The empirical findings demonstrate that the choice of parametric distribution is critical for capturing time-to-event dynamics in infectious disease research. Among all evaluated configurations, models based on the Weibull distribution provided a superior fit to the data, as evidenced by lower AIC values and higher log-likelihoods. This indicates that the assumption of an increasing hazard rate over time, as accommodated by the Weibull shape parameter, represents a statistically valid and clinically realistic framework for modeling survival in acute respiratory infections.

Furthermore, the study highlighted important insights regarding risk factors. Standard AFT models confirmed that advanced age, diabetes, and neurological disorders significantly decrease patient survival time. When accounting for long-term survivors through mixture cure models—which identified a baseline cure fraction of approximately 21% to 27%—age emerged as the primary significant driver of mortality within the susceptible population. The findings suggest that while standard AFT models are highly effective for identifying comprehensive covariate effects across a cohort, mixture cure models provide an essential alternative when a distinct sub-population of cured or non-susceptible individuals exists. Ultimately, selecting the appropriate parametric framework and distribution is essential for generating reliable clinical insights, optimizing prognostic strategies, and informing public health interventions in medical research.

Conflict of Interest

The authors declare there is no existing conflict of interest.

Data Availability Statement

The dataset used in this study was obtained from [20] and is publicly available online. The data can be accessed through the GitHub repository at: https://github.com/gabrielamrodrigues/OLLW/blob/main/data.covid.txt. The dataset was used under the terms specified by the original authors and is available for verification and replication of the study findings.

References

1.Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

2.Bradburn, M. J., Clark, T. G., Love, S. B., and Altman, D. G. (2003). Survival analysis part ii: Multivariate data analysis—an introduction to concepts and methods. *British Journal of Cancer*, 89(3):431–436.

3.Coemans, M., Verbeke, G., Döhler, B., Süsal, C., and Naesens, M. (2022). Bias by censoring for competing events in survival analysis. *BMJ*, 378:e069349.

4.Collett, D. (2023). *Modelling survival data in medical research*. Chapman and Hall/CRC, 4th edition.

5.Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with a cure fraction. *Biometrics*, 38(4):1041–1046.

6.Gaddale, S., Belay, D. M., Gosa, A. M., and Mohammed, S. A. (2023). Predictors of mortality among hospitalized covid-19 patients: A parametric survival analysis approach. *PLoS ONE*, 18(11):e0293962.

7.George, B., Seals, S., and Aban, I. (2014). Survival analysis and log-rank test. *Frontiers in Public Health*, 2:145.

8.Hougaard, P. (1999). Fundamentals of survival data. *International Statistical Review*, 67(2):189–212.

9.Klein, J. P. and Moeschberger, M. L. (2003). Survival analysis: techniques for censored and truncated data. *Biometrics*, 53:1182.

10.Kuitunen, I., Artama, M., Mäkelä, L., Backman, K., Heiskanen-Kosma, T., and Renko, M. (2021). Log-normal distribution provides a good fit for length of hospital stay in children with respiratory tract infections. *Acta Paediatrica*, 110(10):2826–2831.

11.Kuk, A. Y. and Chen, C. H. (1992). A mixture model with proportional hazards for sampled survival data with a cure fraction. *Biometrika*, 79(3):531–541.

12.Lambert, P. C., Thompson, J. R., Weston, C. L., and Dickman, P. W. (2007). Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, 8(3):576–594.

13.Lawless, J. F. (2011). Statistical models and methods for lifetime data. *John Wiley & Sons*, 52.

14.Lee, E. T. and Wang, J. W. (2013). *Statistical methods for survival data analysis*. John Wiley & Sons, 4th edition.

15.Lesko, C. R., Edwards, J. K., Cole, S. R., Moore, R. D., and Lau, B. (2018). When to censor? *American Journal of Epidemiology*, 187(3):623–632.

16.Marazzi, A., Ruffieux, C., and Beguin, C. (1998). Log-normal and weibull distributions for length of stay. *Health Care Management Science*, 1(1):37–47.

17.Othus, M., Barlogie, B., Leblanc, M., and Crowley, J. (2012). Cure models as a useful statistical tool for analyzing survival data. *Clinical Cancer Research*, 18(14):3731–3737.

18.Peng, Y. and Dear, K. B. (2000). A non-parametric mixture cure model for survival data. *Biometrics*, 56(1):237–243.

19.Prinja, S., Gupta, N., and Verma, R. (2010). Censoring in clinical trials: Review of survival analysis techniques. *Indian Journal of Community Medicine*, 35(2):217–221.

20.Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

21.Swindell, W. R. (2009). Accelerated failure time models provide a useful alternative to proportional hazards manager for survival analysis. *Journal of Clinical Epidemiology*, 62(3):245–253.

22.Sy, J. P. and Taylor, J. M. (2000). Estimation in a cox proportional hazards cure model. *Biometrics*, 56(1):227–236.

23.Zare, A., Mahmoodi, M., Mohammad, K., and Zeraati, H. (2014). Comparison of parametric accelerated failure time and cox proportional hazard models in survival analysis of breast cancer patients in iran. *Iranian Journal of Public Health*, 43(11):1543–1550.

© 2026 by the authors. Disclaimer/Publisher's Note: The content in all publications reflects the views, opinions, and data of the respective individual author(s) and contributor(s), and not those of Sphinx Scientific Press (SSP) or the editor(s). SSP and/or the editor(s) explicitly state that they are not liable for any harm to individuals or property arising from the ideas, methods, instructions, or products mentioned in the content.