

# Modeling Survival Time to Recovery from COVID-19: A Case Study on Singapore

**Monira Mollazehi**

Hamad Medical Corporation

**Mohammad Mollazehi**

Qatar University

**Abdel-Salam G. Abdel-Salam** (✉ [abdo@qu.edu.qa](mailto:abdo@qu.edu.qa))

Qatar University <https://orcid.org/0000-0003-4905-6489>

---

## Research article

**Keywords:** Corona Virus, COVID-19, Survival Analysis, Parametric Model, Log-Logistic Regression Model

**Posted Date:** March 30th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-18600/v2>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background** Novel coronavirus (COVID-19) is a new strain of viruses that originated in China. In December 2019, a strange case of pneumonia was reported in Wuhan, which then was diagnosed to be COVID-19. In Singapore, the first positive case was reported on January 23, 2020. **Aim** This study aims to study the recovery time from COVID-19 in Singapore between January 23 and March 13, 2020. **Methods** It's a retrospective study from January 23 until March 13, 2020 for 187 cases with COVID-19 infection. Data of the instances were collected to identify the factors affecting the recovery time from COVID-19. Several parametric models were fitted and the best predictor model was selected using the Bayesian information criterion (BIC). **Results** Out of 187 patients, 96 (51.34%) were cured. The mean ( $\pm$ standard deviation) survival time was  $9.40 \pm 7.17$  days. Based on BIC, the exponential regression model was the weakest and the Weibull model was the best for fitting to data. According to the Weibull model, age (HR=1.01) and nationality (HR=0.76) had significant statistical effects ( $\alpha=0.05$ ) on patient recovery time. **Conclusion** According to the findings of the study, younger patients recover faster compared to elderly patients, and the Singaporean patients cure more quickly than non-Singaporean patients.

## Background

Corona Virus is a new human Betacoronavirus that uses densely glycosylated spike protein to penetrate host cells. The COVID-19 belongs to the same family classification with Nidovirales, viruses that use a nested set of mRNAs to replicate and it further falls under the subfamily of alpha, beta, gamma and delta Co-Vis. The virus that causes COVID-19 belongs to the Betacoronavirus 2B lineage and has a close relationship with SARS species. It is a novel virus since the monoclonal antibodies do not exhibit a high degree of binding to SARS-CoV-2. Replication of the viral RNA occurs when RNA polymerase binds and re-attaches to multiple locations (1,2).

Cases of COVID-19 started in December 2019 when a strange condition was reported in Wuhan, China. This virus has a global mortality rate of 3.4%, which makes it more severe in relation to flu. The elderly who have other pre-existing illnesses are succumbing more to the COVID-19. People with only mild symptoms recover within 3 to 7 days, while those with conditions such as pneumonia or severe diseases take weeks to recover. The recovery percentage of patients, for example, in China, stands at 51%. The recovery percentage rate of COVID-19 is expected to hit 90% (3).

The virus has spread from China to 196 other countries and territories across the globe. From Wuhan, Hubei province, the virus spread to Mainland China, Thailand, Japan, South Korea, Vietnam, Singapore, Italy, Iran, and other countries (4). Singapore was one of the countries that were affected in the early stages of the COVID-19 spreading. The first positive case reported in Singapore was a Chinese national from Wuhan who tested positive for the virus on January 23, 2020 and the number continued to rise to reach up to 509 confirmed cases by March 24, 2020 (5).

This study aims to model the recovery time of patients infected with COVID-19 in Singapore, using several parametric models. Also, to examine the relationship between a patient's recovery time from COVID-19 and the contributing factors of age, gender, and nationality.

## Methodology

In this research, there are 509 cases that tested positive for the novel coronavirus (COVID-19) between January 23<sup>rd</sup> and March 24<sup>th</sup>, 2020. These cases were registered in different hospitals in Singapore. The data has been collected from published records on COVID-19 by the Singapore Government (5) and the Ministry of Health (6) according to the following pattern:

Each patient of the 509 patients is referred to a case number; case 1 up to case 509, and for each case, several factors were reported, such as age, gender, nationality, date case was diagnosed and the date cured or died. For example, the 1<sup>st</sup> case was diagnosed positive as laboratory-confirmed Covid-19 infection on the 23<sup>rd</sup> of January, which was for a 66 years old Chinese male, who was discharged from the hospital (completely cured) on the 19<sup>th</sup> of February 2020. Table 1 shows an example of the dataset used in the current study.

Table 1: Covid-19 Cases Dataset

Case ID	Gender	Age	Nationality	Diagnosed Date	Cured/Death Date	Status
1	Male	66	Chinese	23/1/2020	19/2/2020	Cured
.	.	.	.	.	.	.
.	.	.	.	.	.	.
509	Male	53	Singaporean	23/3/2020	.	.

The study used several parametric models, such as the Exponential, Weibull, Logistic, Log-Logistic, Normal, Log-Normal and Generalized Gamma were tested using SAS software. In order to compare those models and select the best-fitted model, the Bayesian information criterion (BIC) was utilized. The BIC is a measure of the goodness of fit. The model with a smaller value indicates that the proposed model fits the data well (7).

## Results

### 3.1. Patients Characteristic

Out of the 509 patients with COVID-19 included in this study, 29.86% (152) of them were cured while the rest 70.14% (357) were censored. In this case, all censored instances considered to be right-censored, which means they either died or didn't recover and were still undergoing treatment by March 24, 2020. The average time of a patient stay in the hospital was  $7.86 \pm 6.46$  days. In addition, 60.12% (306) of

patients were male, while 39.88% (203) were female. The average age of male patients was 42.86 ( $\pm 16.46$ ) years, while for female patients its was 43.48( $\pm 16.85$ ) years.

Table 2: Demographic characteristics of the COVID-19 cases in Singapore

Characteristics		N (%)	Mean	SD
Age (year)	Male	306 (60.12)	42.86	16.46
	Female	203 (39.88)	43.48	16.85
	Total	509 (100)	43.11	16.60
Time (day)	Male	306 (60.12)	7.83	6.72
	Female	203 (39.88)	7.90	6.07
	Total	509 (100)	7.86	6.46
Age (year)	Singaporean	369 (72.50)	43.47	17.04
	Others	140 (27.50)	42.16	15.42
	Total	509 (100)	43.11	16.60
Time (day)	Singaporean	369 (72.50)	7.48	6.33
	Others	140 (27.50)	8.84	6.71
	Total	509 (100)	7.86	6.46

The BIC criterion for the Normal and Logistic models were 1248.55 and 1239.55, respectively, which are the highest among all parametric models. For the Exponential model, the BIC criterion was 620.31. Moreover, the BIC criterion for the Log-Normal, Generalized Gamma, and Weibull models were 491.32, 489.23, and 489.13, respectively. The Log-Logistic model had the lowest BIC with the value of 481.90, which infer that this model has a better fit to this data (Table 2). Based on the plot of predicted probabilities against recovery time using Log-Logistic distribution, it's clearly inferred that the Log-Logistic regression model had a better fit to the data (Figure 1).

Table 3: Comparison of the BIC between Parametric Models

Model	Number of Parameters	BIC
Exponential	4	620.31
Weibull	5	489.13
Logistic	5	1239.55
<b>Log-Logistic</b>	<b>5</b>	<b>481.90</b>
Normal	5	1248.55
Log-Normal	5	491.32
Generalized Gamma	5	489.23

### 3.2. Log-Logistic Accelerated Failure Time (AFT) Model

(See Log-Logistic Accelerated Failure Time (AFT) Model in the Supplementary Files)

### 3.4. Modeling the Recovery Time using Log-Logistic Distribution

Based on the multiple Log-Logistic regression model, after adjusting other factors, the hazard ratio of age is 1.01 (95% CI: 1.01, 1.01), which is statistically significant at the level of 5%. Similarly, the hazard ratio of nationality was 0.76 and significant (95% CI: 0.64, 0.91). The only insignificant factor was the gender of patients (HR = 0.91, 95% CI: 0.78, 1.07).

Table 4: Diagnostic Factors of the COVID-19 Cases Using Multiple Log-Logistic Regression Model

Parameters	HR	Chi-Square	P-value	95% CI HR	
Intercept	9.26	277.71	0.00	7.13	12.03
Age	1.01	10.66	0.00	1.00	1.01
Gender					
Female	0.91	1.22	0.27	0.78	1.07
Male (Reference)					
Nationality					
Singaporean	0.76	9.21	0.00	0.64	0.91
Non-Singaporean (Reference)					

## Discussion

In modeling survival data, the semiparametric proportional hazards models are utilized frequently. In these models, no particular distributions are

required for the data. Only a critical assumption called the PH assumption is needed to be checked (11). On the other hand, parametric models might perform better. In this study, several parametric models were fitted and compared to each other using the BIC. Among seven parametric models, the Log-Logistic had the lowest BIC, which means that the Log-Logistic model fitted this data better than the other parametric models.

The objective of the study was to discover the factors that affect the recovery time of patients from COVID-19 infection in different hospitals in Singapore. Among the 509 patients that were included in this study, 29.86% were cured before March 24, 2020.

Based on the results of the Log-Logistic model, age and nationality are playing a significant role in survival time to recovery from COVID-19 infection. The results show that the hazard rate of recovery time from COVID-19 infection will increase by 1% as the age of patient increase by one year. In terms of nationality, the hazard rate of Singaporean patients was 0.76 times the Non-Singaporean patients.

The main limitation attributed to the study is that it relies on a limited number of factors, which are gender, age, and nationality. Other demographic information and patients' vital signs were not obtainable due to the unavailability of these records online. A high percentage of the cases were considered as right-censored, and the reason is attributed to the increased number of positive cases before the end of the study. Those cases were still in hospital under treatment at the time when the data collection was completed on March 24, 2020. Further studies for a longer period and a bigger number of cases with more factors included, could help in providing more insight about the recovery time from COVID-19.

## Conclusion

In conclusion, the findings show that younger patients recover faster compared to elderly patients, and local patients cure more quickly than other patients. The study included all the cases diagnosed with COVID-19 in Singapore between January 23<sup>rd</sup> until March 24<sup>th</sup>, 2020. Also, it is the first research that analyzed the factors affecting the recovery time from COVID-19 infection using several parametric models.

## Declarations

**Ethics approval and consent to participate:** Not Applicable.

**Consent for publication:** Not Applicable.

**Availability of data and materials:** The dataset is available from the corresponding author upon request.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article

### **Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

**Authors' contributions:** Authors participated in drafting the article, and they gave final approval of the version to be submitted.

**Acknowledgments:** Thanks for the reviewers and the Editor.

## **References**

1. McIntosh K. Coronavirus disease 2019 (COVID-19). UpToDate. <https://www.uptodate.com/contents/coronaviruses>. Accessed March 10 2020
2. Fisher D, Heymann D (2020) Q&A: The novel coronavirus outbreak causing COVID-19. BMC Medicine. doi: 10.1186/s12916-020-01533-w
3. World Health Organization. Coronavirus. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed March 12 2020
4. Coronavirus Cases. <https://www.worldometers.info/coronavirus/#countries>. Accessed March 24 2020
5. Singapore Government. Coronavirus disease 2019: Cases in Singapore. <https://www.gov.sg/article/covid-19-cases-in-singapore>. Accessed March 24 2020
6. Ministry of Health. Updates on COVID-19 (Coronavirus Disease 2019) Local Situation. <https://www.moh.gov.sg/covid-19>. Accessed March 11 2020
7. Weakliem DL. A Critique of the Bayesian Information Criterion for Model Selection. *Sociological Methods & Research*. 1999; 27:359–397
8. Faruk A. The comparison of proportional hazards and accelerated failure time models in analyzing the first birth interval survival data. *Journal of Physics: Conference Series*. 2018; doi:10.1088/1742-6596/974/1/012008
9. Swindell WR. Accelerated failure time models provide a useful statistical framework for aging research. *Experimental Gerontology*. 2009;44(3):190–200.
10. Liu X. *Survival analysis: models and applications*. Chichester: A John Wiley & Sons; 2012.

## Figures

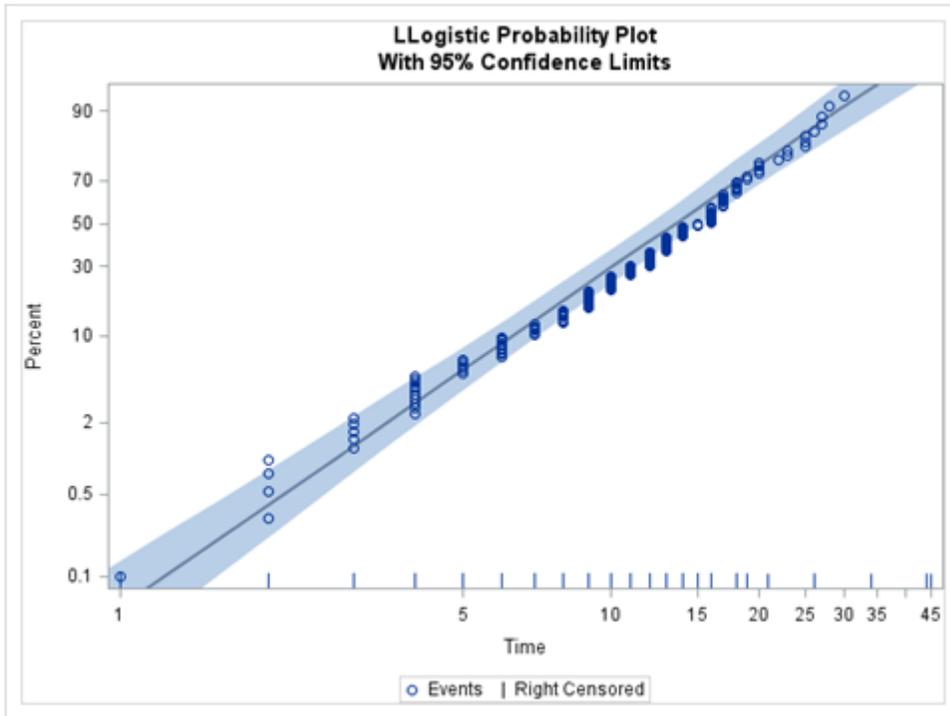


Figure 1

Log-Logistic Probability Plot with 95% Confidence Limits for Recovery Time

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [LogLogisticAcceleratedFailureTimeAFTModel.pdf](#)