Retrospective Clinical Research Report

Journal of INTERNATIONAL MEDICAL RESEARCH

Journal of International Medical Research
49(8) 1–7
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/03000605211040263
journals.sagepub.com/home/imr
SAGE

Modeling the recovery time of patients with coronavirus disease 2019 using an accelerated failure time model

Gayathri Thiruvengadam[1], Ravanan Ramanujam[2] and Lakshmi Marappa[3]

Abstract

Objective: To identify factors associated with recovery time from coronavirus disease 2019 (COVID-19).

Methods: In this retrospective study, data for patients with COVID-19 were obtained between 21 June and 30 August 2020. An accelerated failure time (AFT) model was used to identify covariates associated with recovery time (days from hospital admission to discharge). AFT models with different distributions (exponential, log-normal, Weibull, generalized gamma, and log-logistic) were generated. Akaike's information criterion (AIC) was used to identify the most suitable model.

Results: A total of 730 patients with COVID-19 were included (92.5% recovered and 7.5% censored). Based on its low AIC value, the log-logistic AFT model was the best fit for the data. The covariates affecting length of hospital stay were oxygen saturation, lactate dehydrogenase, neutrophil-lymphocyte ratio, D-dimer, ferritin, creatinine, total leucocyte count, age > 80 years, and coronary artery disease.

Conclusions: The log-logistic AFT model accurately described the recovery time of patients with COVID-19.

[1]Faculty of Allied Health Sciences, Sri Ramachandra Institute of Higher Education and Research, Porur, Chennai, India
[2]Collegiate Education, Chennai Region, Chennai, India
[3]Department of General Medicine, Sri Ramachandra Institute of Higher Education and Research, Porur, Chennai, India

Corresponding author:
T. Gayathri, Faculty of Allied Health Sciences, Sri Ramachandra Institute of Higher Education and Research, Porur, Chennai 600116, India.
Email: gayathri.t@sriramachandra.edu.in

Keywords
Accelerated failure time model, Akaike's information criterion, coronavirus disease-19, length of stay, recovery time, modeling

Date received: 21 May 2021; accepted: 29 July 2021

Introduction

The World Health Organization (WHO) has described the 2000s as the "century of non-communicable diseases". Globally, the incidence of non-communicable diseases such as type 2 diabetes mellitus, hypertension, and dyslipidemia is increasing, resulting in increased morbidity and mortality. At the end of 2019, a virulent communicable disease called coronavirus disease 2019 (COVID-19) began to spread and subsequently developed into a global pandemic. COVID-19, caused by severe acute respiratory syndrome coronavirus-2, has affected every part of the world and is associated with high morbidity and a large economic burden. Managing COVID-19 is a major challenge for healthcare workers. As a communicable disease, COVID-19 affects all ages and sexes. Symptoms vary from mild and moderate to severe or fatal in some patients. Mild symptoms include pyrexia, sore throat, and dry cough. Some patients also report loss of taste and smell, severe headache, myalgia, joint pain, and gastrointestinal symptoms. Fewer patients exhibit severe symptoms such as chest pain and shortness of breath. The WHO has reported that approximately 85% of patients exhibit mild to moderate symptoms and recover well. By contrast, 10% to 15% of patients exhibit severe symptoms, and 5% require intensive care unit support with high mortality rates. A small proportion of individuals who contract COVID-19 remain asymptomatic and can spread the infection to others. Globally as of April 2021, 134 million cases of COVID-19 had been reported; approximately 107 million individuals recovered and approximately 2 million died. In India, 12.9 million cases of COVID-19 were reported; 11.8 million individuals recovered and 166,892 deaths were reported by the Union Ministry of Health and Family Welfare.[1] Recovery rate and recovery time among patients with COVID-19 depend upon various factors. The WHO has stated that the mean recovery time for patients with mild to moderate symptoms is around 14 days. By contrast, recovery requires around 3 to 6 weeks in individuals with severe COVID-19. A meta-analysis showed that the mean recovery time from symptom onset was 22.2 days (range: 18–83 days).[2]

A study of 538 Chinese patients with COVID-19 (January to March 2020) found that the mean duration of hospital stay was 19 days (interquartile range 14–23 days). Recovery time depended on age, COVID-19 severity, and area of residence.[3] Another study of Belgian patients with COVID-19 reported hospitalization durations of 3 to 10.4 days after accounting for factors such as age, sex, and nursing home admission.[4] A study of 221 patients with COVID-19 in India reported an average recovery time of 25 days (95% confidence interval 16–34 days); this estimate accounted for only sex and age.[5]

Accurate identification of the recovery time and length of hospital stay of patients with COVID-19 is an urgent priority for healthcare delivery. Recovery time also depends upon the underlying health status of the individual including any comorbid conditions such as diabetes, hypertension, cardiopulmonary disease, and renal failure.

In this study we also examined laboratory parameters such as creatinine, lactate dehydrogenase (LDH), ferritin, hemoglobin (Hb), total leukocyte count (TC), polymorphonuclear leukocytes, lymphocytes, and D-dimer. These factors were not assessed in previous studies.

Methodology

This was a retrospective cohort study. The study was approved by the institutional ethics committee of Sri Ramachandra Institute of Higher Education and Research (EC-NI/20/AUG/75/51). Because this was a retrospective study of medical records, the requirement for informed consent was waived. The study was registered in the Clinical Trials Registry – India. Patients with COVID-19 admitted to the hospital of Sri Ramachandra Institute of Higher Education and Research were included. Patients admitted for other reasons and diagnosed with COVID-19 as part of routine evaluation were not included. Patient data were collected from medical records including demographic characteristics, past and current medical history (e.g., metabolic and other comorbid diseases), clinical symptoms of COVID-19 (e.g., fever, sore throat, cough, myalgia, loss of taste and smell, and headache), and laboratory parameters such as oxygen saturation (SpO_2), creatinine, LDH, ferritin, Hb, TC, neutrophil-lymphocyte ratio (NLR), and D-dimer.

The semiparametric Cox model is the most common statistical tool used for time to event analysis as it clearly defines the effects of covariates on the outcome variable. The parametric accelerated failure time (AFT) model is a better choice only if the distribution of survival time is well defined. In the AFT model, interpretation is simple because it is directly dependent on survival time; in this model the effects of covariates are directly proportional to survival time. In the current study, we examined the recovery time from COVID-19 and factors influencing it among patients in a tertiary care hospital in Chennai using an AFT model. Recovery time (length of hospital stay) was the dependent variable. Patients who died or were discharged against medical advice were censored.

AFT model

In survival analysis, various statistical procedures can be used to identify factors associated with response variables over time. Parametric models are classified into proportional hazards models and AFT models. AFT models use log-logistic, log-normal, and generalized gamma distributions whereas proportional hazards models use the Gompertz distribution. Weibull and exponential distributions are used in both proportional hazards and AFT models.

The AFT model is formulated as:

$$\mathrm{S(t)} = \mathrm{S_0}(\varphi \mathrm{t}),$$

where φ is the acceleration factor expressed as $\exp(b_1x_1 + b_2x_2 + \cdots + b_px_p)$ and (x_1, $x_2, \ldots, x_p$) are the covariates.

The time ratio (TR) is calculated by taking the exponent of the coefficient (exp (b_i)). A positive coefficient (b_i) or TR >1 indicates that the effect of the covariate prolongs recovery time, whereas a negative coefficient or TR < 1 indicates a shorter recovery time compared with the reference category.

To select the most suitable model, Akaike's information criterion (AIC) was calculated as follows:

$$\mathrm{AIC} = -2 * \mathrm{Log\ likelihood} + 2 * (\mathrm{a} + \mathrm{c}),$$

where *a* represents the number of parameters in the model and *c* denotes the constant coefficients depending upon the distribution.

For Weibull, log-logistic, log-normal, and generalized gamma distributions the constant coefficient is 2, while for the exponential distributions the constant coefficient is 1.[6]

Statistical analysis

Data were entered in MS Excel (Microsoft, Redmond, WA, USA) and analyzed using STATA software version 14.0 (StataCorp, College Station, TX, USA). Univariate analysis was performed for all variables separately. Covariates showing statistical significance in the univariate analysis were included in the multivariate AFT model. The AIC was used to select the best-fitting AFT model (exponential, log-normal, generalized gamma, log-logistic, or Weibull distribution). Values of $p < 0.05$ were considered statistically significant.

Results

This was a retrospective cohort study comprising 730 patients with COVID-19 (456 men and 274 women). Among the participants, 675 (92.5%) recovered, 30 (4.1%) were discharged against medical advice, and 25 (3.4%) died. Thus, 55 (7.5%) patients were censored. The median age was 48 years (interquartile range 35–61 years). Almost all patients (97.4%) had COVID-19 symptoms. The most common symptom at the time of hospital admission was fever (86.9%) followed by sore throat (50.5%) and cough (38.8%). The least common symptom at the time of hospital admission was vomiting (1.4%) and loose stool (2.5%). The median recovery time of hospitalized patients was 7 days. The mean duration from symptom onset to hospitalization was 4.6 days (standard deviation 2.3 days).

In AFT models using exponential, Weibull, generalized gamma, log-normal, and log-logistic distributions, 2, 5, 8, 8, and 9 parameters were significantly associated with time to recovery, respectively. Irrespective of model, SpO_2 and LDH were significantly associated with recovery time. Other than in the model using an exponential distribution, NLR was significantly associated with recovery time for all other distributions. As shown in Table 1, the most suitable distribution was log-logistic as shown by its low AIC value (AIC = 675.398). Using the AFT model with log-logistic distribution, covariates significantly associated with recovery time were SpO_2, LDH, NLR, creatinine, D-dimer, ferritin, TC, age >80 years, and coronary artery disease. Clinical symptoms, sex, and other comorbid conditions including diabetes, hypertension, and chronic kidney disease were not significantly associated with recovery time. The AFT model with log-logistic distribution showed that age >80 years vs. ≤60 years (TR = 1.461, $p = 0.017$), SpO_2 <95% vs. ≥95% (TR = 1.481, $p < 0.001$), ferritin <25 ng/mL or >200 ng/mL vs. 25–200 ng/mL (TR = 1.129, $p = 0.017$), TC <4000 cells/µL or >11,000 cells/µL vs. 4000–10,000 cells/µL (TR = 1.111, $p = 0.007$), NLR >3 vs. ≤3 (TR = 1.098, $p = 0.01$), D-dimer >0.55 µg/mL vs. ≤0.55 µg/mL (TR = 1.094, $p = 0.01$), LDH >300 international units (IU)/L vs. ≤300 IU/L (TR = 1.181, $p < 0.001$), creatinine >1.4 mg/dL vs. ≤1.4 mg/dL (TR = 1.226, $p = 0.038$), and coronary artery disease (TR = 1.268, $p = 0.004$) were associated with prolonged length of hospitalization. The goodness of fit of the AFT model was tested. The AFT model with log-logistic distribution had a R^2 value of 0.47, indicating that 47% of variation in recovery time was explained by the model.

Discussion

The current study aimed to identify factors associated with length of hospital stay among patients with COVID-19. The most commonly used statistical tool for analyzing time to event data is the Cox proportional hazards model.[7] An AFT model is a

Table 1. Parameters of multivariate AFT model of time to recovery from COVID-19.

Covariates		Exponential			Weibull			Log−logistic			Log−normal			Generalized gamma		
		ß	SE	P-value	ß	SE	P-value	ß	SE	P-value	ß	SE	P-value	ß	SE	P-value
Sex	Male	0.005	0.094	0.957	0.043	0.048	0.364	−0.026	0.035	0.446	−0.014	0.037	0.708	−0.016	0.035	0.649
	Female (ref)															
Age	≤60 years (ref)															
	61–80 years	0.045	0.11	0.679	−0.022	0.056	0.693	0.038	0.04	0.338	0.016	0.042	0.703	−0.003	0.04	0.94
	>80 years	0.707	0.425	0.098	0.316	0.215	0.142	0.321	0.134	0.017	0.361	0.139	0.009	0.378	0.124	0.002
Cough	Yes	−0.009	0.120	0.934	−0.031	0.061	0.615	0.002	0.044	0.969	0.007	0.048	0.869	0.0245	0.047	0.6
	No (ref)															
Breathlessness	Yes	−0.07	0.143	0.642	−0.044	0.074	0.556	0.021	0.052	0.687	−0.002	0.056	0.962	−0.005	0.055	0.928
	No (ref)															
Sore throat	Yes	−0.021	0.114	0.849	−0.005	0.058	0.936	−0.007	0.042	0.868	−0.028	0.045	0.54	−0.046	0.043	0.288
	No (ref)															
Vomiting	Yes	0.235	0.405	0.562	0.248	0.204	0.225	0.314	0.158	0.059	0.261	0.154	0.089	0.245	0.143	0.087
	No (ref)															
Loss of smell or taste	Yes	0.027	0.204	0.896	−0.082	0.104	0.428	−0.048	0.073	0.507	0.028	0.081	0.732	0.048	0.077	0.533
	No (ref)															
Creatinine (mg/dL)	≤1.4 (ref)															
	>1.4	0.176	0.264	0.504	0.105	0.133	0.425	0.204	0.098	0.038	0.192	0.106	0.07	0.243	0.107	0.024
D-dimer (μg/mL)	≤0.55 (ref)															
	>0.55	0.122	0.094	0.193	0.083	0.048	0.083	0.09	0.035	0.011	0.085	0.037	0.02	0.079	0.034	0.023
NLR	≤3 (ref)															
	>3	0.159	0.1	0.111	0.145	0.053	0.006	0.094	0.037	0.011	0.108	0.038	0.005	0.09	0.036	0.012
SpO_2 (%)	≥95 (ref)															
	<95	0.466	0.152	0.002	0.406	0.079	<0.001	0.398	0.057	<0.001	0.393	0.06	<0.001	0.372	0.059	<0.001
Ferritin (ng/mL)	25–200 (ref)															
	<25 or >200	0.122	0.136	0.372	0.186	0.069	0.007	0.121	0.051	0.016	0.106	0.054	0.049	0.084	0.052	0.103
LDH (IU/L)	≤300 (ref)															
	>300	0.228	0.107	0.031	0.205	0.056	<0.001	0.166	0.039	<0.001	0.178	0.041	<0.001	0.178	0.038	<0.001
TC (cells/mL)	4000–11,000 (ref)															
	<4000 or >11,000	0.125	0.108	0.251	0.078	0.055	0.158	0.106	0.04	0.007	0.096	0.042	0.021	0.084	0.039	0.033
DM	Yes	0.082	0.097	0.398	−0.001	0.051	0.975	0.056	0.035	0.105	0.068	0.038	0.075	0.081	0.035	0.021
	No (ref)															
HT	Yes	−0.024	0.107	0.82	−0.001	0.055	0.983	−0.007	0.0391	0.857	−0.023	0.042	0.568	−0.029	0.039	0.46

(continued)

Table 1. Continued.

Covariates		Exponential			Weibull			Log−logistic			Log−normal			Generalized gamma		
		ß	SE	P-value	ß	SE	P-value	ß	SE	P-value	ß	SE	P-value	ß	SE	P-value
	No (ref)															
CKD	Yes	−0.151	0.308	0.623	−0.079	0.154	0.608	−0.032	0.125	0.796	−0.033	0.126	0.792	−0.023	0.128	0.858
	No (ref)															
CAD	Yes	0.38	0.208	0.067	0.31	0.104	0.003	0.237	0.083	0.004	0.206	0.077	0.007	0.133	0.073	0.068
	No (ref)															
Constant		0.306	0.423	0.47	0.6	0.217	0.006	0.451	0.46	0.004	0.507	0.168	0.003	0.492	0.163	0.003
Log likelihood		−691.072			−477.16			−315.95			−333.44			−319.18		
AIC		1422.14			996.32			675.398			708.88			682.37		

AFT, accelerated failure time; COVID-19, coronavirus disease 2019; SE, standard error; ref, reference; NLR, neutrophil to lymphocyte ratio; SpO_2, oxygen saturation; LDH, lactate dehydrogenase; IU, international units; TC, total leukocyte count; DM, diabetes mellitus; HT, hypertension; CKD, chronic kidney disease; CAD, coronary artery disease; AIC, Akaike's information criterion.

parametric survival model used to assess the effects of covariates on a response variable, and is an alternative to the Cox model for time-to-event data. AFT models do not require Cox proportional hazards modeling as an *a priori* test.[8] For parametric models, it is necessary to identify the appropriate distribution for survival time. Identifying the appropriate distribution will result in better fit of the model to the data. AFT model interpretation directly relates to the survival time and time ratio rather than the hazard ratio. For time to event analysis, Cox models are very common and only a few studies of breast cancer, gastric cancer, ovarian cancer, anxiety disorder, and liver failure have used AFT models. A recent study of patients with COVID-19 conducted in Singapore found that an AFT model with log-logistic distribution was more suitable than other AFT models.[9] This is in agreement with the results of our study, which showed that the AFT with log-logistic distribution was the most suitable model. This model showed that abnormal values of SpO_2, creatinine, LDH, NLR, D-dimer, ferritin, TC, age >80 years, and coronary artery disease were associated with longer hospital stays. These results are consistent with recent studies of D-dimer levels[10] and the NLR.[10,11] However, our findings regarding LDH disagree with the results of a study conducted in China,[10] which found that LDH was not associated with length of hospital stay but was instead associated with increased risk of complications.

In the current study, COVID-19 symptoms were not associated with length of hospital stay. In contrast, a study conducted in China reported an association between dyspnea and length of hospitalization.[10] Our results showed that sex was not associated with length of hospitalization, in agreement with the results of previous studies.[4,5,10] Thus, the AFT model with log logistic distribution is the most suitable model for predicting time to recovery among patients with COVID-19.

Limitations

This was a retrospective cohort study conducted among 730 patients with COVID-19 in a single tertiary care hospital in southern India. Hence, future studies with larger sample sizes should be conducted among individuals of various ethnicities to extend the results of the current study.

Conclusions

An AFT model with log-logistic distribution was the most suitable model for predicting length of hospital stay among patients with COVID-19. Clinical parameters such as SpO_2, creatinine, LDH, NLR, D-dimer, ferritin, TC, age >80 years, and coronary artery disease were associated with time to recovery. Understanding the factors that reduce length of hospitalization could help in developing strategies to optimize delivery of health care services.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of conflicting interest

All the authors have no conflicts of interest.

Author contributions

Study design: Gayathri T, Ravanan R.
Data collection: Gayathri T, Lakshmi M.
Statistical Analysis: Gayathri T.
Results interpretation: Gayathri T, Ravanan R.
Manuscript preparation: Gayathri T, Ravanan R, Lakshmi M.

ORCID iD

Gayathri Thiruvengadam https://orcid.org/0000-0003-2753-0717

References

1. Worldometer. COVID-19 coronavirus pandemic. Coronavirus update (live). https://www.worldometers.info/coronavirus (accessed 1 February 2021).
2. Rees EM, Nightingale ES, Jafari Y, et al. COVID-19 length of hospital stay: A systematic review and data synthesis. *BMC Med* 2020; 18: 270.
3. Wang Z, Ji JS, Liu Y et al. Survival analysis of hospital length of stay of novel coronavirus (COVID-19) patients in Sichuan, China. *medRxiv* 2020. doi: https://doi.org/10.1101/2020.04.07.20057299.
4. Faes C, Abrams S, Van Beckhoven D, et al. Time between symptom onset, hospitalisation and recovery or death: Statistical analysis of Belgian COVID-19 patients. *Int J Environ Res Public Health* 2020; 17: 7560.
5. Barman MP, Rahman T, Bora K, et al. COVID-19 pandemic and its recovery time of patients in India: A pilot study. *Diabetes Metab Syndr* 2020; 14:1205–1211.
6. Dietz K, Gail M, Krickeberg K, et al. *Statistics for biology and health*. New York: Springer, 2000, p. 549.
7. Thiruvengadam G, Lakshmi M and Ramanujam R. A study of factors affecting the length of hospital stay of COVID-19 patients by Cox-proportional hazard model in a South Indian tertiary care hospital. *J Prim Care Community Health* 2021; 12: 21501327211000231.
8. Khanal SP, Sreenivas V and Acharya SK. Accelerated failure time models: An application in the survival of acute liver failure patients in India. *Int J Sci Res (Ahmedabad)* 2020; 3:161–166.
9. Mollazehi M, Mollazehi M and Abdel-Salam AS. Modeling survival time to recovery from COVID-19: A case study on Singapore. *Research Square* 2020; https://www.researchsquare.com/article/rs-18600/v2 (accessed 20 November 2021).
10. Chen FJ, Li FR, Zheng JZ, et al. Factors associated with duration of hospital stay and complications in patients with COVID-19. *J Public Health Emerg* 2021; 5: 6.
11. Yang A, Liu J, Tao W, et al. The diagnostic and predictive role of NLR, d-NLR and PLR in COVID-19 patients. *Int Immunopharmacol* 2020; 84: 106504.