

# Competing risk between in-hospital mortality and recovery: An application of DeepHit on COVID-19 clinical data

Lintu M.K.

Manipal Academy of Higher Education <https://orcid.org/0000-0002-7693-8615>

Asha Kamath (✉ [asha.kamath@manipal.edu](mailto:asha.kamath@manipal.edu))

Manipal Academy of Higher Education <https://orcid.org/0000-0003-0727-8067>

---

## Research Article

**Keywords:** competing risk, COVID-19, DeepHit, in-hospital mortality, survival analysis

**Posted Date:** February 8th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-103866/v2>

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

[Read Full License](#)

---

# Abstract

## Background

Unexplained pneumonia appeared in Wuhan was soon determined to be a novel coronavirus disease, referred to as COVID-19. On March 11, 2020, WHO characterized COVID-19 as a pandemic. A plethora of studies on this pandemic is being carried out using various statistical and mathematical models. Though most of them are focusing on building predictive models, concentrating on the length of hospital stay can improve decision making and treatment plans. While modeling the length of stay, possible outcomes observed are either discharge or mortality.

## Objective

The study aimed to analyse the survival data of COVID-19 patients with the competing risk methodology.

## Methodology

The performance DeepHit, a deep learning-based competing risk model is compared with Fine-Gray, a traditional statistical model using a time-dependent concordance index.

## Results

The deep learning-based competing risk model outperformed the statistical model in terms of discriminative power.

## Conclusion

Modeling the duration of recovery and death provides valuable information for health officials to design proper strategies during the outbreak. These outcomes should be considered as competing events to model the data adequately.

# Introduction

The coronavirus disease 2019(COVID-19) outbreak is unfortunately still under progression. The virus has affected most countries across the world. The severity of the disease is associated with a large number of co-morbidities. From the previous studies, it has been found out that gender, age, and pre-existing chronic diseases such as cardiovascular disease, cancer, respiratory disease, etc. are associated with increased risk for adverse outcomes.<sup>1,2</sup> In addition to this information, it is important to focus on the length of hospital stay and multiple outcomes of the COVID-19 patients. Because of the lack of clinical data, not many studies have come out with these objectives.

Survival analysis considers the problem of analyzing data where the target variable is time until the occurrence of an event. Events could be anything such as death, hospitalization, hospital discharge, etc. The challenge underlying the scenario is to learn the model parameters from survival time data while handling censoring. The concept of censoring comes into the picture since the survival time is unknown for some of the subjects due to lost to follow-up or culmination of the study period.<sup>3,4</sup>

The common objective of lifetime data analysis is understanding the relationship between the covariates and time-to-event.<sup>5</sup> The traditional format of survival data consists of a tuple, for each of the subjects. is the time until the occurrence of the first event or censoring, and indicates the censoring status, and is the covariates. While dealing with time-to-event data, it is common to observe more than one outcome occurring among the subjects, which is referred to as competing events. Competing risk setting is ubiquitous in epidemiological studies and clinical trials since the subjects are likely to experience multiple possible adverse events. Since the problem is challenging to address, most of the existing models treat one of the events as the event of interest and others as right censoring.<sup>6</sup> This approach is inadequate as the occurrence of an event of interest is often obscured by other competing events.<sup>7</sup> Fine-Gray model is one of the survival techniques which has looked into this competing risk aspect. Hence, it is considered as a benchmark in competing risk analysis.<sup>8</sup> With the increasing acceptance of deep learning models, deep learning-based competing risk models are emerging as an important predictive tool in medicine. These models attempt to give empirical estimates of the true cumulative incidence functions.<sup>9-14</sup> DeepHit is one such deep-learning technique developed to handle the competing events with good empirical performance.<sup>6,14</sup> We explored the use of these models to analyse the survival characteristics of 1863 COVID-19 patients admitted to hospitals across the globe from December 10, 2019, to March 30, 2020. In this data, discharge, and death being two events, we framed the competing risks setup.<sup>15-17</sup> Implementing this kind of predictive models help clinicians to know the possible outcome of the patients in advance which aid in improving treatment policies during the outbreaks or epidemics.

In the rest of the paper, we present the data, explain the methods, and discuss the results.

## Methods

The performance of a deep learning-based competing risk model(DeepHit) on a COVID-19 clinical data is compared with a traditional statistical model(Fine-Gray). These are the two benchmarks for the competing risk problem in the current literature. Both of these models are based on the cumulative incidence function(CIF). Cause-specific CIF gives the probability that the event occurs on or before time conditional on the covariates . i.e.,

$$F(t, k|x) = P(T \leq t, k|x)$$

## Fine-Gray model

Fine-Gray<sup>8</sup> model is the most commonly used statistical method in competing risk problems. The traditional proportional hazards model is modified by the direct transformation of the CIF. This approach focus on the subdistribution of a competing risk. For each failure type, the model provides a direct interpretation in terms of survival probabilities. The *cmprsk* package in R is used to fit the Fine-Gray model. We used the materials provided by Nemchenko et al.<sup>7</sup> to fit the model and find the performance metric.

## DeepHit

DeepHit<sup>6</sup> is a deep neural network that learns the distribution of survival times directly without making any assumption on the underlying stochastic process. The model trains a multi-task network to learn the estimate of the joint distribution of the first hitting time and competing events. DeepHit can be used to predict the competing risks, discharge from the hospital (event 1), and death prior to discharge (event 2). Since we are considering 2 competing events, the network consists of 4-layers. The first layer is a fully-connected layer for the shared subnetwork, followed by two fully-connected layers for each cause-specific sub-network. The output layer is a softmax layer. ReLU(Rectified Linear Unit) activations are used in all three layers. The network training is done by back-propagation via Adam optimizer with a batch size of 50 and a learning rate of . The dropout probability of 0.1 and Xavier initialization was applied for all the layers. We used the *pycox* package to implement the model.<sup>13</sup> For training, testing, and validation, 60%, 20%, and 20% of the data are randomly separated. For evaluation, 5-fold cross-validation is being applied.

## Evaluation metric

The time-dependent concordance index is used to evaluate the discriminative ability of the models.<sup>18</sup> The principle of concordance is that the predicted survival probability of a subject who experienced the event should be less than those who have survived longer. The value ranges between 0 and 1. As the metric approaches one, better the performance of the model. The concordance index evaluates a method's discriminative performance.

## Data

The raw data of 1863 hospitalized patients is extracted from an open-access COVID-19 epidemiological data website [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(20\)30119-5/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30119-5/fulltext).

A group of researchers collected epidemiological data from different research labs. The data was extracted from online resources and national health portals released by state/local health officials and hospitals of different countries. The dataset consists of a subject ID, date of hospital admission, gender, age, the onset date of symptoms, outcome(death/discharge), death/discharge date, history of chronic disease, symptoms, location, and travel history.

Time to the events is calculated directly by subtracting the hospital admission date from the date of the outcome. Observations without the admission dates, covariate information, and outcome were removed from the study. Discharge from the hospital and in-hospital mortality is event 1 and event 2 respectively. Censoring time is obtained from the last available date. Since the outcomes were well defined, there was no complication in defining death, discharge, and censoring. We have included only the patients who got admitted till March 30, 2020, to avoid massive censoring. We have considered the covariates: age, gender, chronic disease history, latitude, and longitude for the analysis. Table 1 presents outcome-wise descriptive statistics. All the percentages are calculated based on the remainder of 1863. The schematic plot for competing-risk time-to-event data for five hypothetical subjects can be seen in Fig. 1.

Table 1. Event-wise summary measures of the patients.

Descriptive statistics	Event 1(Discharge)	Event 2(Death)
No. of subjects(%)	162(8.7)	59(3.2)
Median time to event in days(Range)	13(8-18)	7.5(4-10.75)
Median age in years(Range)	38(30-51)	70(62.75-79.75)
No. of males(%)	94(5.1)	39(2.1)

## Results

The time-to-event data of 1863 COVID-19 patients were analysed. Among these, 162(8.7%) discharges and 59(3.2%) deaths were reported. The socio-demographic descriptions of the events in the COVID-19 cases are depicted in Fig. 2. Mortality was marginally higher among the males(Fig. 2A.), higher among the elderly (Fig. 2B.), and among those with pre-existing chronic diseases(Fig. 2C.). Whereas, the discharge was marginally higher among males, higher among age group and among those without pre-existing chronic disease condition.

CIF curves obtained for the events are depicted in Fig. 3. The estimated joint distribution of survival time with two competing events is trained with DeepHit. Fine-Gray model encodes the linear effects of patient covariates on both the events.

The average discriminative indexes of both the models are given in Table 2. DeepHit outperformed the Fine-Gray model, indicates the better discriminative power of DeepHit. Due to the small number of observations(3.2% of the entire dataset), the performance of the models for event 2 is variable compared to event 1. This happens when a lesser number of event 2 is being sampled in the validation and testing sets.

Table 2. Average cause-specific index.

Model	Mean (95% CI)	
	Event 1 (Discharge)	Event 2 (Death)
Fine-Gray	0.731(0.711,0.751)	0.690(0.672,0.708)
DeepHit	0.872(0.868,0.876)	0.800(0.788,0.812)

## Discussion And Conclusions

As several studies pointed out, we also observed that elders and patients with chronic diseases are more prone to die than other age groups and the gender and age of the patients have a direct effect on their recovery time.<sup>19-21</sup> Few studies have also reported that being old or male, the probability of hospital discharge is lower.<sup>22</sup>

Cox proportional hazards model has been used to study the mortality and recovery of COVID-19 patients. The bias in estimating the hazard ratio of death due to COVID-19 in the presence of competing events has been investigated.<sup>23</sup> These studies point out the need of considering recovery and death as competing events to avoid a substantial risk of misleading results.<sup>20,24</sup>

The upper hand of DeepHit over the other competing risk models has already been established in many studies.<sup>6,14</sup>

Due to the lack of availability of clinical data, competing risk modeling of COVID-19 data are less explored to the best of our knowledge.<sup>25</sup> This is the first study to use a deep learning-based competing risk model on COVID-19 clinical data with discharge and death being two competing events. There is a huge scope for further analysis when detailed data with other potential risk factors are available. An interesting expansion of the study includes the development of more sophisticated deep learning algorithms to measure the impact of clinical and prognostic factors on potential survival.

## Declarations

### Conflicts of interest/Competing interests

The authors declare that there is no conflict of interest.

### Funding

Not applicable.

### Availability of data and material

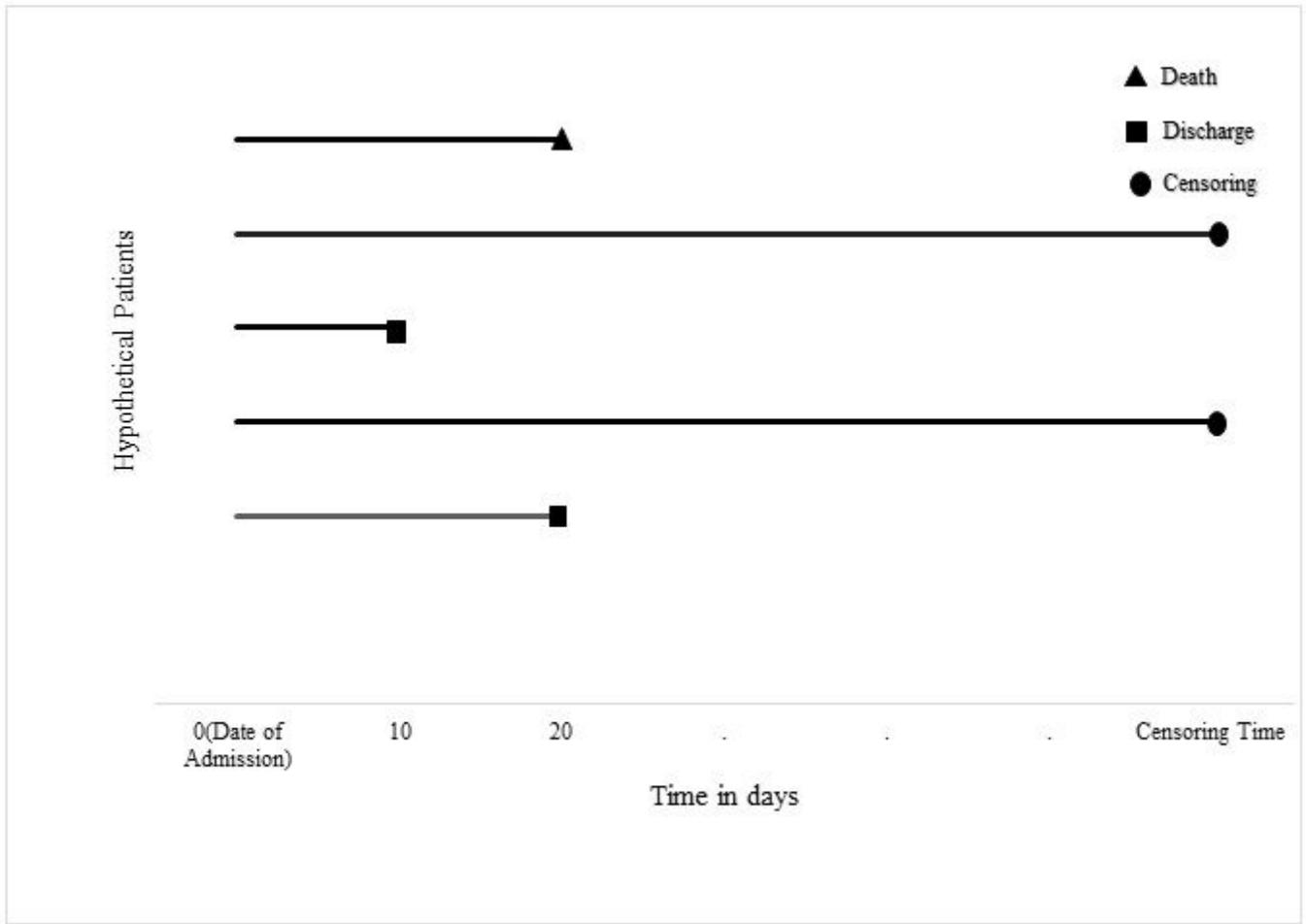
Data is available from open-access COVID-19 epidemiological data website [https://www.thelancet.com/journals/laninf/article/PIIS1473-3099\(20\)30119-5/fulltext](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30119-5/fulltext).

# References

1. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *Jama*. 2020 Apr 7;323(13):1239-42.
2. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KS, Lau EH, Wong JY, Xing X. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*. 2020 Jan 29.
3. Hosmer Jr DW, Lemeshow S, May S. *Applied survival analysis: regression modeling of time-to-event data*. John Wiley & Sons; 2011 Sep 23.
4. Kleinbaum DG, Klein M. *Survival analysis*. Springer;2010.
5. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1972 Jan;34(2):187-202.
6. Lee C, Zame WR, Yoon J, van der Schaar M. DeepHit: A Deep Learning Approach to Survival Analysis With Competing Risks. In *AAAI 2018 Apr 26* (pp. 2314-2321).
7. Nemchenko A, Kyono T, Van Der Schaar M. Siamese Survival Analysis with Competing Risks. In *International Conference on Artificial Neural Networks 2018 Oct 4* (pp. 260-269). Springer, Cham.
8. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*. 1999 Jun 1;94(446):496-509.
9. Al-Shedivat M, Dubey A, Xing EP. Personalized survival prediction with contextual explanation networks. *arXiv preprint arXiv:1801.09810*. 2018 Jan 30.
10. Faraggi D, Simon R. A neural network model for survival data. *Statistics in medicine*. 1995 Jan 15;14(1):73-82.
11. Bellot A, Schaar M. Tree-based bayesian mixture model for competing risks. In *International Conference on Artificial Intelligence and Statistics 2018 Mar 31* (pp. 910-918).
12. Gupta G, Sunder V, Prasad R, Shroff G. CRESA: A Deep Learning Approach to Competing Risks, Recurrent Event Survival Analysis. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining 2019 Apr 14* (pp. 108-122). Springer, Cham.
13. Kvamme H, Borgan Ø, Scheel I. Time-to-event prediction with neural networks and Cox regression. *Journal of machine learning research*. 2019;20(129):1-30.
14. Lee C, Yoon J, Van Der Schaar M. Dynamic-deephit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE Transactions on Biomedical Engineering*. 2019 Apr 3;67(1):122-33.
15. Awad A, Bader–El–Den M, McNicholas J. Patient length of stay and mortality prediction: a survey. *Health services management research*. 2017 May;30(2):105-20.
16. Oulhaj A, Ahmed LA, Prattes J, Suliman A, Al Suwaidi A, Al-Rifai RH, Sourij H, Van Keilegom I. The competing risk between in-hospital mortality and recovery: A pitfall in COVID-19 survival analysis research. *medRxiv*. 2020 Jan 1.

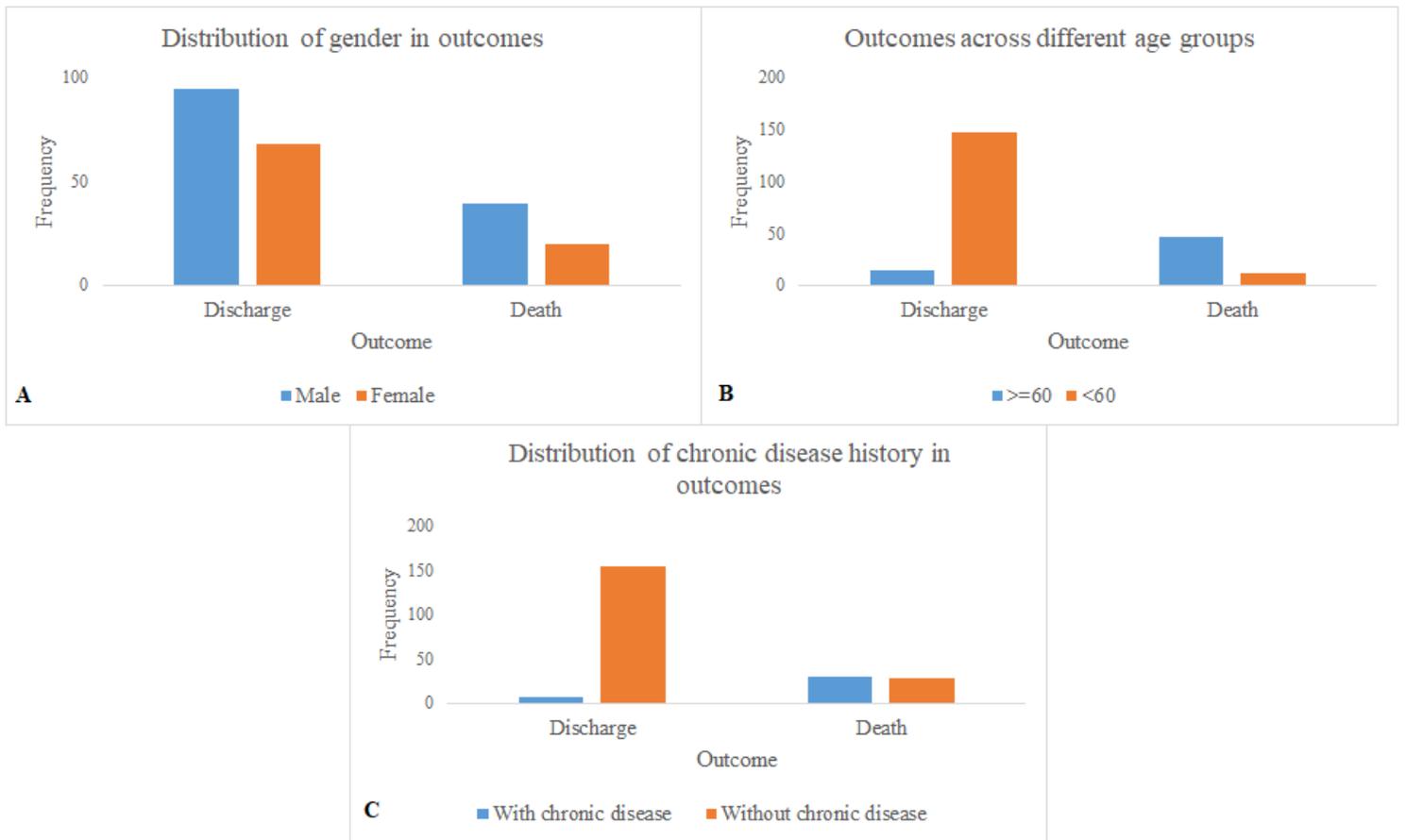
17. Lu M, Ishwaran H. Dynamic Competing Risk Modeling COVID-19 in a Pandemic Scenario. *emergence*. 2020 Apr 8.
18. Antolini L, Boracchi P, Biganzoli E. A time-dependent discrimination index for survival data. *Statistics in medicine*. 2005 Dec 30;24(24):3927-44.
19. Jordan RE, Adab P, Cheng KK. Covid-19: risk factors for severe disease and death.
20. Li X, Xu S, Yu M, Wang K, Tao Y, Zhou Y, Shi J, Zhou M, Wu B, Yang Z, Zhang C. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. *Journal of Allergy and Clinical Immunology*. 2020 Apr 12.
21. Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, Li Q, Jiang C, Zhou Y, Liu S, Ye C. Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. *Journal of Infection*. 2020 Apr 23.
22. Nemati M, Ansary J, Nemati N. Machine-learning approaches in COVID-19 survival analysis and discharge-time likelihood prediction using clinical data. *Patterns*. 2020 Aug 14;1(5):100074.
23. Li LQ, Huang T, Wang YQ, Wang ZP, Liang Y, Huang TB, Zhang HY, Sun W, Wang Y. COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis. *Journal of medical virology*. 2020 Jun;92(6):577-83.
24. Ghosh S, Samanta GP, Mubayi A. COVID-19: Regression Approaches of Survival Data in the Presence of Competing Risks: An Application to COVID-19. *Letters in Biomathematics*. 2020 May 7.
25. Wolkewitz M, Lambert J, von Cube M, Bugiera L, Grodd M, Hazard D, White N, Barnett A, Kaier K. Statistical analysis of clinical covid-19 data: A concise overview of lessons learned, common errors and how to avoid them. *Clinical epidemiology*. 2020;12:925.

## Figures



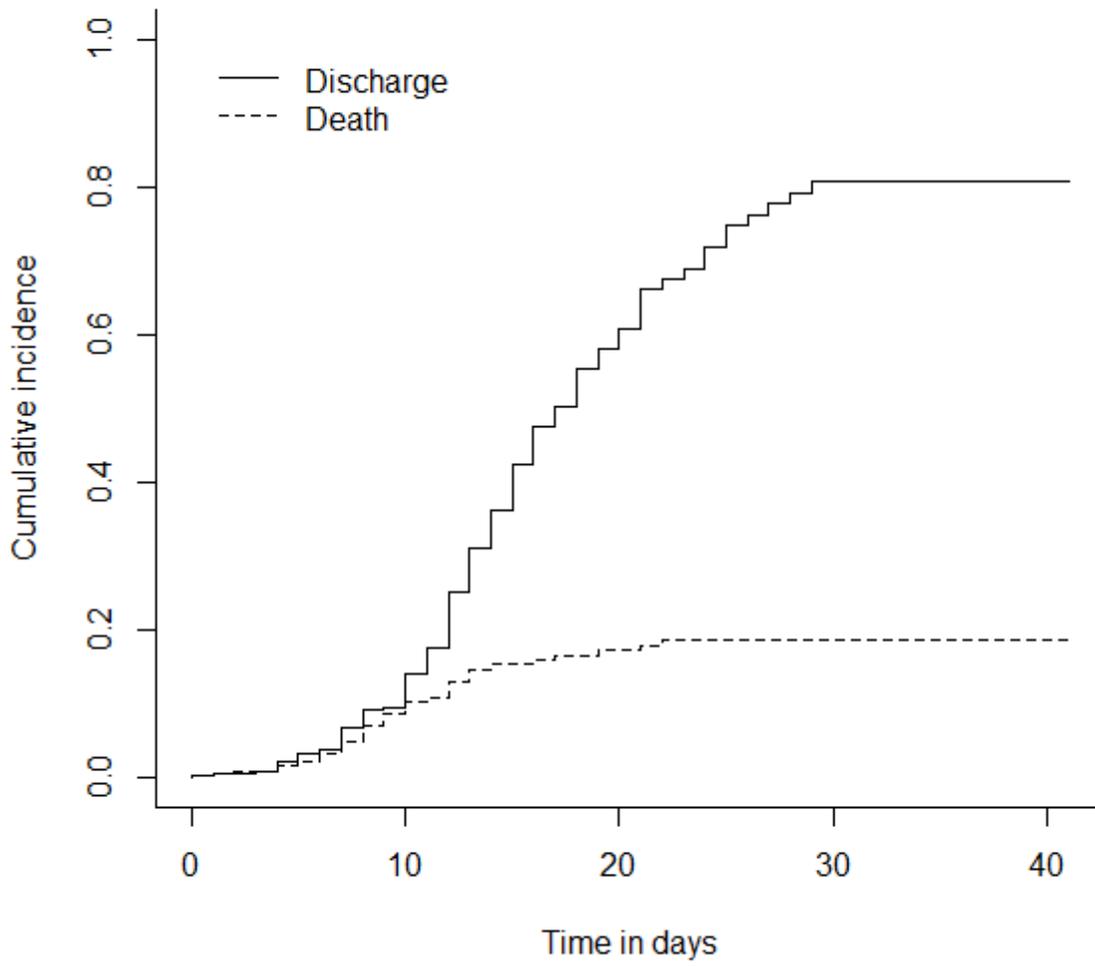
**Figure 1**

Schematic plot for competing-risk time-to-event data for five hypothetical subjects.



**Figure 2**

The socio-demographic descriptions of the outcomes in the COVID-19 patients.



**Figure 3**

Cumulative incidence functions.