Survival Analyses of COVID-19 Patients in a Turkish Cohort: Comparison between Using Time to Death and Time to Release

Sirin Cetin [1*], Ayse Ulgen [2*], Pervin Ozlem Balci [3], Hakan Sivgin [4], Meryem Cetin [5], Sevdiye Sivgin [6], Wentian Li [7*]

[1] *Department of Biostatistics, Faculty of Medicine, Tokat Gaziosmanpaşa University, Turkey.*

[2] *Department of Biostatistics, Faculty of Medicine, Girne American University, Karmi, Cyprus.*

[3] *Department of Medical Microbiology, Tokat State Hospital, Tokat, Turkey.*

[4] *Department of Internal Medicine, Tokat State Hospital, Tokat, Turkey.*

[5] *Department of Medical Microbiology, Faculty of Medicine, Tokat Gaziosmanpaşa University, Turkey.*

[6] *Department of Anesthesiology and Reanimation, Tokat State Hospital, Tokat, Turkey.*

[7] *The Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, United States.*

Abstract

Survival analyses of COVID-19 data has its own unique features, in particular, the existence of two distinct events: death and release from the hospital within a very short period of time. This multiple-event situation belongs to a type where the occurrence of the first event prevents the second event to happen, and vice versa. We carried out two cause-specific univariate Cox regression survival analyses, one for time-to-death and another for time-to-release. Each survival analysis is further split into one for onset of symptom to event time and another for hospitalization to event time. We have also carried out a case-control (death vs. release) analysis without considering the time to event information. We observed that risk factors can be detected by either case-control or survival analysis, even though the goal of the two is quite different. We also observed that the two survival analyses may not both reveal a factor being a risk factor, but only one of them does. We prefer this two rounds of Cox regressions over mixture cure model which is only focused on time-to-death events which usually are sample size limited. By utilizing time-to-release events may greatly increase the sample size needed for revealing risk factors for COVID-19.

Keywords: COVID-19; Survival Analysis; Cause-specific Hazard Ratio; Case-control Analysis.

1. Introduction

COVID-19 is a new infectious disease causing global pandemic on a scale not seen in one hundred years, since the 1918 flu pandemic. Much has been learned about the disease in the past year: it is caused by the coronavirus SARS-Cov-2 (previously called 2019-nCov), a single-stranded RNA virus of 30kb genome size with around a dozen genes [1]; the disease is unusually infectious, perhaps made even more so after mutations such as an aspartic acid (D) to glycine (G) mutation at position 614 amino acids of the spike protein gene [2, 3], an asparagine (N) to tyrosine (Y)

mutation at position 501 in B.1.1.7 variant ("UK variant") [4], and glutamine (E) to lysing (K) mutation at position 484 for several variants including B.1.351 ("South Africa variant") [5], P.1 ("Brazil/Japan variant") [6], B.1.526 ("New York variant") [7]; though COVID-19 is a respiratory disease targeting primarily lung, the virus invades epithelial cells in blood vessels and other organs and it is far from a single-organ disease.

Epidemiology studies on COVID-19 are extensive. Older age is a top risk factor for death from COVID-19. Comorbidities such as obesity and diabetes are also dominant risk factors [8-10]. Most of the studies are static in the sense that association between a factor and an outcome is tested disregard of any time factor. A severe COVID-19 patient may recover quickly from a treatment or die after a long battle in ICU despite all efforts by the doctors. The time the patient spent in hospital towards either total recovery or death is an extra piece of information to reveal how a factor conveys risk.

We all know that time-to-event-outcome is studied in survival analysis. In regression survival analysis (e.g. Cox proportional-hazard model), the dependent variable is the time from a baseline point to the event, plus a status value dealing with right censored samples where status=0 if it is unknown whether the event occurs or not. The best example of survival analysis in medical research is cancer survival: event is cancer caused death, and time is the time lapse from the day of (e.g.) diagnosis to the day of death. If a patient is alive or quits the study, the time is the time lapse from the day of diagnosis to the day at the end of the study (or the day when the patient quits); and his/her status is zero.

In order to apply survival analysis to COVID-19 patient data, we have to deal with new situations. First, unlike cancer where there is only one kind of event, i.e., death (being alive is considered as having an unknown status towards the eventual death), there are two kinds of event for COVID-19: death and released from the hospital. Second, cancer survival time is often measured by years whereas COVID-19 patients stay in hospital for a much shorter time. It is quite likely that there are no right-censored samples in the dataset, and all patients are either released/cured or died. How do we handle two time to the events: time to death and time to release?

In Allison (2014) [11] study, five situations are discussed when there are multiple kinds of events (I, IIa, IIb, IIc, IId). Let us assume there are only two kinds of events, and the dependent variable is $y = Surv(time, status)$ where status=1 (first kind of event),2 (second kind of event), or 0 (right censored). Situation-I: one mechanism determines the time value, whereas another mechanism determines if status is 1 or 2. Situation-IIa: if a person has a status=1 event, the person no longer has the risk to have the status=2 event (and vice versa). Situation-IIb: if a person has a status=1 event, the person is no longer available for observation of a status=2 event (and vice versa). Situation-IIc: the risk for two kind of events are independent. Situation-IId: if a person has a status=1 event, the risk for the status=2 event is changed (and vice versa).

Obviously, if a COVID-19 patient dies, he/she would not be released as a healthy person, and if a patient is released we consider his COVID-19 disease is cured and will no longer die from COVID-19. These observations put COVID-19 time-to-event data under the situation-IIb group. And the recommended steps is to carry about two rounds of survival analysis, with the first round treating event-1 as status=1 and event-2 as status=0, and the second round switching the two [11].

A portion of the field of competing-risk survival analysis is dominated by the study of correct estimation of survival curve under multiple events where some event (e.g. surgical operation) contributes to the survival/nonsurvival status of a patient. Without considering the contribution from the new type of event, the Kaplan-Meier estimation of the survival curve would be biased. In particular, there is a subdistribution (sd) method by Fine & Gray (1999) [12]. In Fine-Gray approach, the estimated of hazard ratio (sdHR) is guaranteed in the same direction of the relative risk (i.e., $sdNR > 1 \leftrightarrow RR > 1$) – to have a refresher on the definition of both [13]. Some authors support the use of Fine-Gray method in COVID-19 data [14], and others actually use it [15].

However, not only the interpretation of sdHR can be confusing (e.g., P Allison (2018), "For causal analysis of competing risks, don't use Fine & Gray's subdistribution method", *http://statisticalhorizons.com/for-causal-analysis-of-competing-risks*), but also the release and death events are not one preceding another, but mutually exclusive, a situation not compatible with the situation for applying the Fine-Gray method. Many papers have explored the relationship between cause-specific hazard ratio (csHR), which is resulted from our practice described above as Allison's situation IIb, and sdHR, e.g. [16-19]. There are following conclusions: even if the proportional hazard ratio assumption is violated, it is still ok to calculate csHR, and csHR correctly measures the effect of a factor on hazard for a sample who has not yet experienced any one of the two events. If the goal of a survival analysis is to establish association with a factor and not to estimate the survival curve, csHR is a better choice than sdHR.

We carry out this two-event survival analysis (i.e., two rounds of survival analysis by using either time-to-death or either-to-release) to estimate cause-specific hazard ratio and the corresponding p-values on a Turkish hospital COVID-19 patients. Intuitively, we may expect the two rounds of survival analyses would detect the same risk factors but in opposite directions: the presence of a risk factor value may shorten the time-to-death, but the absence of the risk value

may also shorten the time-to-release. Our comparative analysis will show that the real situation is more complicated. We also carry out "static" analysis or case-control analysis by ignoring the time-to-event information. The case samples being the deceased patients and control samples are the released ones. Intuitively, we may not expect case-control analysis and survival analysis would lead to the same conclusion as the questions asked are different. However, we find the results from the two analyses are actually more close to each other than expected.

2. Data and Methods

Patient data: Patients (n=109) over 18 years of age who were admitted with COVID-19 in the Tokat State Hospital were included. Patients were treated on COVID-19 regular wards in COVID-19 services, intermediate care and intensive care units (ICU) run by different departments. All patients had a laboratory confirmed case of COVID-19 as defined by a positive result on a reverse-transcriptase-polymerase-chain-reaction (RT-PCR) test of nasopharyngeal swab. Exclusions consisted of patients who left the hospital against medical advice and those who transfered to external hospitals. We record the symptoms prompting presentation to the event time, and time of hospitalization days to release or death.

Admission data included demographics such as gender, age, and job type. We did not use the job type because of small sample sizes per category. Comorbidities were recorded in the following eight categories: lung disease (COPD or other chronic pulmonary disease), heart disease (coronary artery disease/ischemic cardiomyopathy or heart failure NYHA II-IV), diabetes mellitus, chronic liver disease active malignancy, primary or secondary immunodeficiency, obesity (body mass index (BMI) > 30kg/m2), and hypertension (two nurse administered measures by mercury blood pressure monitor with both systolic/diastolic blood pressure higher than 140/90). Each person can have one, multiple, or no comorbidities. We did not use cancer, liver disease, and immunodeficiency because of the small sample size per comorbidity group. Other patient's information include tobacco history, and if the patient is a health care worker.

We have the following common laboratory blood test measures: complete blood count (CBC) with absolute lymphocyte count (including red blood cell count, white blood cell count, platelet count, lymphocyte count, lymphocyte percentage, neutrophil count, neutrophil percentage, monocyte count, eosinophil count), metabolic chemistry panel (BMP/CMP) (including albumin), C-reactive protein (CRP), fibrinogene, D-dimer and other proteins. Laboratory data depicted the first test result occurring within the first 24 hours of hospitalization.

Some patient's information has obvious association with mortality and these are not included as factors, including pneumonia status, degree of pneumonia, being in intensive care unit (ICU), being on incubation machine.

Programs used: All analyses were carried out by R (*www.r-project.org*) packages including *survvial* (functions: coxph, surfit), and the standard statistical functions (fisher.test, wilcox.test, glm(... family="binomial")). The survival analysis with competing risks by Fine-Gray subdistribution function model [12] is carried out by the cmprsk R package.

3. Results

General description of the time-to-events and the analysis tasks: For each patient, we have the time to event (either death or release) from two different starting points: the time of onset of symptom and the time to be admitted to hospital. The cumulative distribution of the time-to-event for four situations is shown in Fig.0. We can see from Fig.0 that although symptom-to-event time is longer than hospitalization-to-event time, the difference between the two is small. Generally speaking the to-event time is longer for death events than for release events. The longest to-event time is one deceased patients who stayed in hospital for 94 days.

We ran three different analyses on the COVID-19 patients data, with distinct goals. The first survival analysis, by Cox regression, on time-to-death data is to find factors where the patient dies earlier when its value increases (positive coefficient) or decrease (negative coefficient). The second survival analysis, also by Cox regression, on time-to-release data is to find factors where the patient release late when its value increases (negative coefficient) or decrease (positive coefficient). The third analysis is the "static" case-control analysis where patients' risk factor is compared between the deceased group and the recovered group. For binary factors, we use the Fisher's test, and for continuous factors, we use the t-test and Wilcoxon non-parametric test. We have added a logistic regression for both categorical (qualitative) and continuous (quantitative) factors.

All our test results (p-values) of three different types are shown in Table 1. For survival analysis, there are two sets of results, one for onset of symptoms to event time, and another for hospitalization to event time. We have added a sign in Table 1 to indicate the sign of Cox coefficients (or whether the hazard ratio is less or larger than 1).
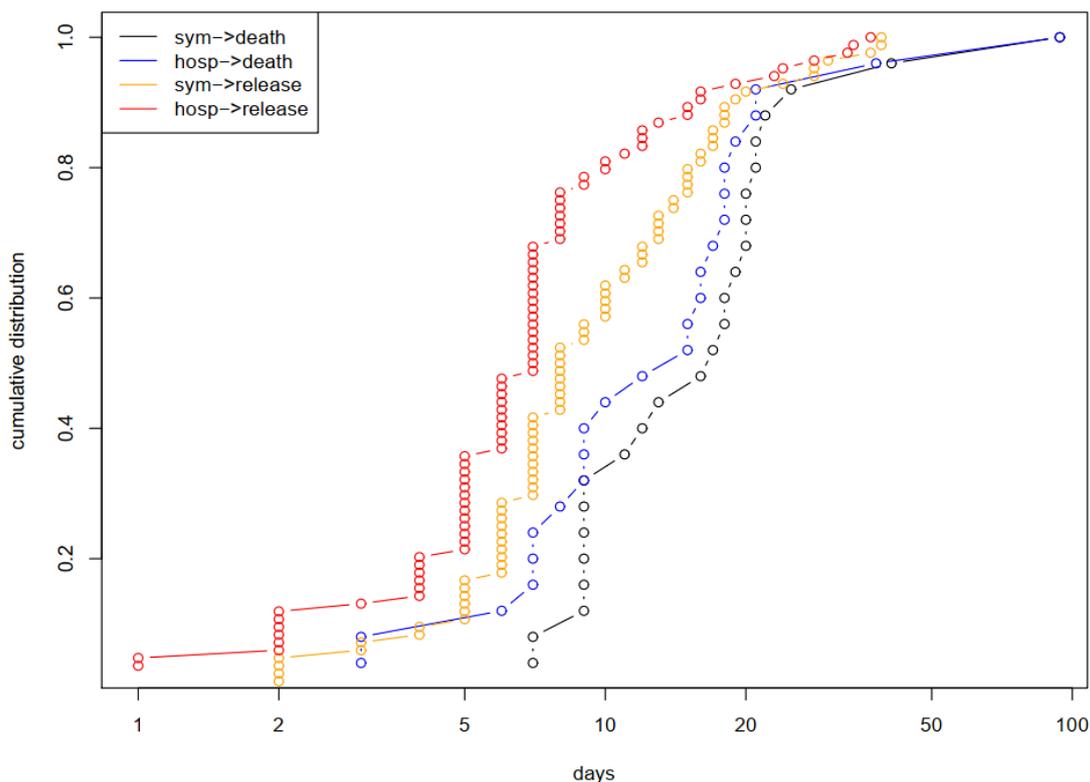
description of various survival times

Figure 1. Cumulative distribution of four time-to-event values: symptom-to-release (orange), hospitalization-to-release (red), symptom-to-death (black), and hospitalization-to-death (blue). The number of released patients is 84 and the number of deceased patients is 25. The y-axis has a jump of 1/84 (for control group) or 1/25 (for case group) for each sample at its time-to-event value (x-axis).

Risk factors detected by static (case-control) analysis are also risk factors in one of the survival analyses, and vice versa: Table 1 shows that if a factor is significant in the static case-control analysis (we tentatively mark those p-values smaller than 0.005 in boldface, following a recommendation in Ioannidis (2018) [20] study, though another choice is to use 0.001 threshold [21], most likely, it will also show up as being significant in one of the survival analyses. The seemingly exceptions to this, monocyte, eosinophil, is only due to a specific choice of the p-value threshold: the p-values from static and temporal analysis are actually comparable.

On the other hand, if a factor is significant in one of the two survival analyses, it is also (most of the time, depending on the meaning of "significance") significant in the case-control analysis. The seemingly exception is hypertension – but again, the p-values in the two situations are comparable.

Opposite direction of a risk factor in two survival analyses: Intuitively, (higher value in) a factor making a patient die earlier should make a surviving patient to stay in hospital longer. In other words, log-hazard-ratio for a given factor should have opposite signs. This can be confirmed by the results in Table 1, where the sign attached to a p-value is the sign for log-hazard-ratio. Whenever there seem to be an exception, e.g., smoking, we found that the p-value does not indicate a statistically significant result. We then expect the hazard ratio fluctuates around the value of 1, and the sign for log-hazard-ratio is either positive or negative by chance.

We may have only one of the two survival analyses produces statistically significant results: By intuition, we would expect the same factor would both significantly make deceased patients die earlier and significantly make surviving patients release later. However, it is more likely by the result in Table 1 that only one survival analysis leads to statistical significant result, whereas another survival analysis does not.

There could be two explanations. The first is the sample size difference between the deceased and survived samples. There are 84 patients who were released from the hospital, whereas there are only 25 patients who died. As contribution to the Cox regression fitting is mainly from samples with the event, it is likely that we see more statistically significant results for time-to-release analysis than that from time-to-dead analysis.

Another possibility is that a risk factor only makes a deceased patient die faster, or makes a survived patient to stay in hospital longer, but not both. Fig.1 shows the survival curves for patients with and without a comorbidity (diabetes, hypertension, ischemic heart disease, obstructive pulmonary disease, and either one of them), in time-to-death and time-to-release survival analysis. For diabetes the survival curve on the left does not give any visual impression that it consistently affects the time-to-disease value, and it does not seem to be due to small sample size. On the other hand, for ischemic heart disease, and for having any one of the four diseases, the time-to-disease survival analysis does show a trend, and that trend could become significant if the sample size is larger.

Table 1. P-values from three types of analyses (time-to-death survival analysis, time-to-release survival analysis, case-control analysis) and 6-7 tests (four Cox regressions, Fisher test for discrete factors, t-test/Wilcoxon-test for continuous factors, logistic regression) for various factors for COVID-19 patients. P-values smaller than 0.005 are marked with boldface. The sign attached is defined as following: for Cox regression, sign of the log(hazard ratio), for t-test/Wilcoxon test, sign of mean(death group) - mean(release group), for Fisher's test, sign of log(odds-ratio) preferring high value in the death group, and for logistic regression, the sign of the regression coefficient. The fabrino and albumin factors are marked by asterisk because they have a very high missing data rate (54% and 67%).

Test p-values (with a sign) of risk factor's association with COVID-19

| factor	time-to-death		time-to-release		"statis"		
	from-sym	from-hosp	from-sym	from-hosp	Fisher	t/Wilcoxon test	logistic Reg
gender(M1/F2)	-0.241	-0.293	0.178	0.132	-0.066		-0.063
age	0.0848	0.246	**-8.6E-5**	**-6.9E-7**		**5.1E-5/3E-4**	**1E-4**
smoking	0.827	0.859	0.836	0.824	-1		-0.89
health worker	-0.709	0.913	0.766	0.245	-1		-0.579
D-dimer	**0.0038**	0.0094	-0.087	-0.029		0.005/ **0.0045**	**0.0029**
CRP	0.0053	0.0066	**-7.3E-5**	**-1.3E-4**		**1.8E-7/9E-7**	**3.6E-5**
fibrinogen *	0.0051	**0.0048**	-0.016	-0.011		**1.6E-4/4.2E-4**	**0.0019**
albumin *	-0.574	-0.89	0.175	0.085		-0.187/0.014	-0.15
RBC	-0.0079	-0.026	**1.7E-7**	**1.3E-9**		**-6.6E-6/2.8E-7**	**-1.3E-5**
hemoglobin	-0.086	-0.18	**8.3E-5**	**5.4E-6**		**-1.1E-4/4.7E-5**	**-9.5E-5**
WBC	**1.6E-4**	**0.001**	-0.0012	-8.2E-4		0.0019/7.5E-4	**2.2E-4**
lymphocyte%	**-3.6E-4**	**-9.8E-4**	**2E-4**	-2.4E-5		**-3.3E-7/7.1E-9**	**-5.2E-6**
lymphocyte	**-9.7E-4**	**0.0022**	0.083	0.043		**-2.1E-6/1E-8**	**-2.9E-5**
neutrophil%	**4.0E-5**	**1.1E-4**	**-1.2E-4**	**-2.1E-5**		**8.7E-9/9.4E-10**	**1.1E-6**
neutrophil	**2.1E-6**	**2.3E-5**	**-3.0E-4**	**-1.3E-4**		**9.3E-5/6.2E-6**	**1.4E-5**
platelet	-0.067	-0.063	-0.497	-0.588		-0.353/0.487	-0.334
monocyte	-0.031	-0.029	0.089	0.11		**-0.0023**/0.0065	-0.016
eosinophil	-0.013	-0.015	0.69	0.62		**-8.4E-5/0.0021**	-0.025
diabetes	0.638	0.902	-0.034	-0.0058	0.046		0.031
hypertension	0.289	0.444	-0.013	**-0.0022**	0.02		0.014
heart disease	0.033	0.069	**-0.0034**	**-0.0014**	**7E-5**		**7E-5**
lung disease	0.511	0.607	**-0.0026**	**-0.0031**	**0.0016**		**0.0014**
above 4 diseases	0.167	0.284	**-1.7E-5**	**-1.2E-6**	**1.8E-4**		**5.2E-4**
kidney disease	0.0496	0.109	-0.099	-0.043	0.028		0.025
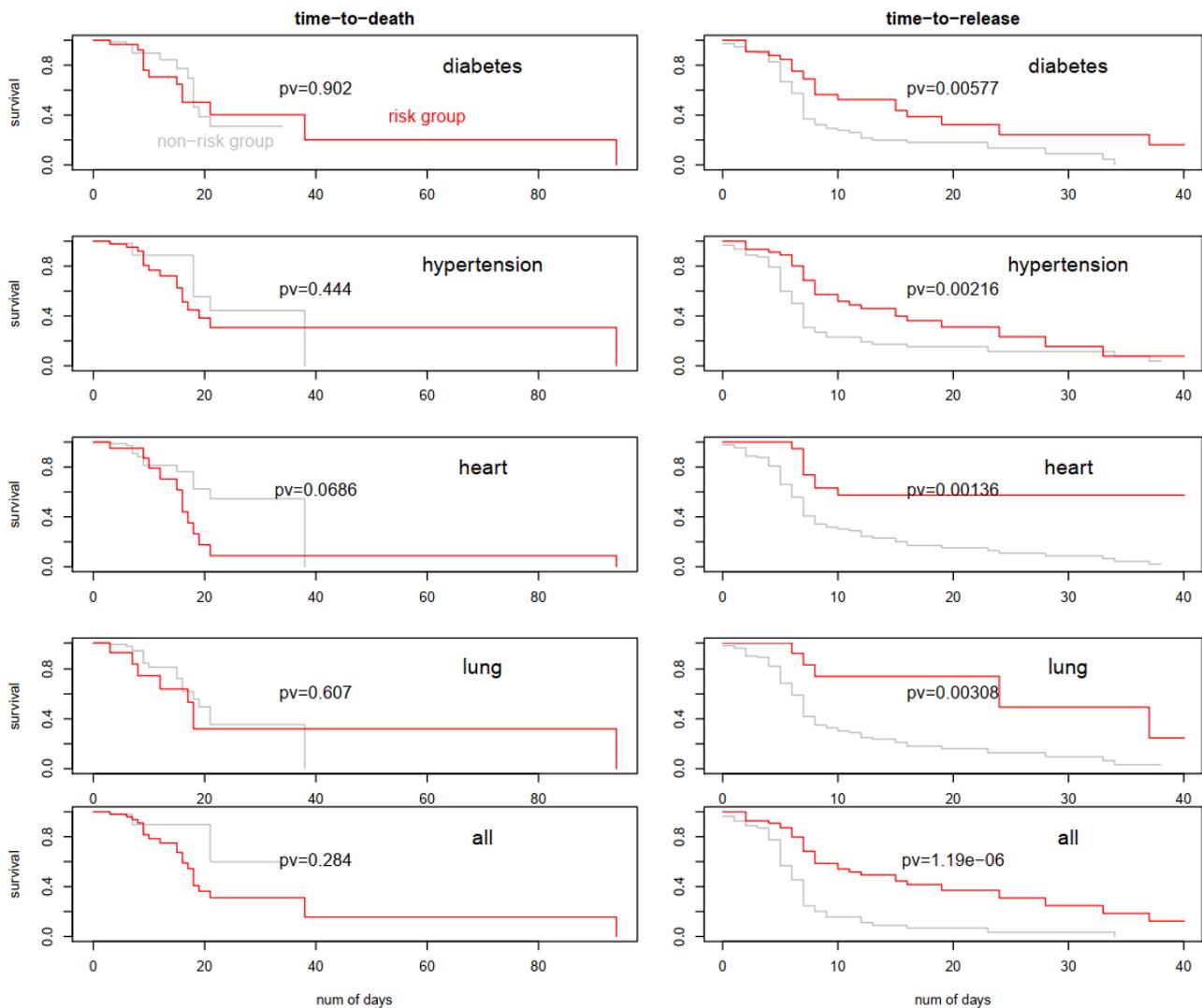
Figure 2. Comparison between the survival curves from time-to-death (left) and those from time-to-release (right) with 4 different comorbidities and the indicating variable for any one of these conditions. Red lines are for those with a comorbidity and grey lines are for those who do not have a comorbidity condition. "Heart" refers to ischemic heart disease, "lung" refers to obstructive pulmonary disease, and "all" refers to the presence of any one of the four comorbidity conditions.

4. Discussion

Since the COVID-19 pandemic in early 2020, there have been many studies to investigate the possible risk factors for infection, severity of the disease, and mortality. Many of these analyses are straightforward case-control analysis [22], i.e., compare the outcome of people with and without certain risk factor value (discrete factors), or its scale (continuous factors). There are also some survival analysis for time-to-death in severe patients [23-28], time-to-release [29], and length of hospital stay [30].

In this paper, we explicitly run three different analyses which address different questions. Case-control analysis asks which factor contributing to a person's outcome, time-to-death survival analysis asks a factor's contribution to a faster death, and time-to-release survival analysis switches the role played by risk factor and protection factor, aiming at finding protection factor that reduce the hospital stay time before release or finding risk factors that lengthens the time. Since these three analyses address different questions, it is not automatic that they should lead to the same results.

The comparison of results from our three different analyses (Table 1) shows that a factor can be a risk in all three different senses, but other factors can be a proven risk only in one or two senses. Our multiple analyses also utilize the dataset more effectively: one may have less number of deceased patients but more number of cured and released patients; then by using the time-to-release survival analysis, more information on factors can be extracted.

Although we prefer the use of cause-specific hazard ratio over sdHR, we still wonder what conclusions the Fine-Gray model leads Table 2 shows the p-values from survival analysis of Fine-Gray model of competing risks for both symptom-to-event and hospitalization-to-event time, for discrete factors only. Comparing Tables 1 and 2, it can be

seen that gender remains to be a weak risk factor, smoking and health worker status are still not risks. Lack of comorbidities remain to be a protection for released patients. Interestingly, comorbidities are not significant risk factors for deceased patients, but two out of five become significant in Fine-Gray model. From Fig.1, we can indeed see slight differences in survival curves between the with and without a comorbidity groups. To summarize all these observation, we conclude that even if the Fine-Gray's approach is used, the conclusion may still be similar to the cause-specific HR approach.

Table 2. The p-values of single-variable competing-risk survival analysis by Fine-Gray model (discrete/categorical factors only)

Test p-values of discrete risk factor's association with COVID-19 by Fine-Gray model

	time-to-death		time-to-release	
factor	from-symptom	from-hosp	from-symptom	from-hosp
gender(M1/F2)	0.073	0.072	0.055	0.044
smoking	0.90	0.89	0.65	0.76
health worker	0.84	0.97	0.72	0.45
diabetes	0.11	0.11	0.035	0.010
hypertension	0.022	0.025	**0.0038**	**0.0012**
heart disease	**5.6E-5**	**6.5E-5**	**0.0011**	**4.7E-4**
lung disease	**0.0025**	**0.0023**	**0.0016**	**0.0021**
above 4 diseases	**0.0015**	**0.0031**	**9.4E-6**	**2E-6**
kidney disease	0.0051	0.0055	0.029	0.027

An alternative analysis framework is the mixture cure model where the released patients are considered as cured with a probability [31-33]. There are several arguments caution on the use of this approach: first, the survival curve of all samples has to be flat in the large time-to-event value, but all of the time-to-event values in our data are small, making it difficult to see the plateau; second, our sample size is too small to fit a model with more parameters; third, the deceased samples, whose number is rather small in our data, play a big role in the mixture cure model. Our cause-specific Cox regression can take advantage of the larger number of released patients to find potential risk factors.

5. Conclusion

In conclusion, we investigated two questions in a COVID-19 survival data. One question is whether survival analysis (considering the time to the event) can detect risk factors not detected by case-control analysis (not considering the time to the event). Another question is time-to-death survival analysis and time-to-release survival analysis may lead to different lists of risk factors. The tentative answers from our n=109 sample size to the above two questions are no and "mostly no but sometimes yes". Datasets with larger sample sizes are needed to make a more conclusive statement.

6. Declarations

6.1. Author Contributions

Conceptualization, S.C., A.U. and W.L.; methodology, S.C., A.U. and W.L.; formal analysis, S.C., A.U. and W.L.; data curation, P.O.B., H.S., M.C. and S.S.; writing—original draft preparation, S.C., A.U. and W.L.; writing—review and editing, P.O.B., H.S., M.C. and S.S. All authors have read and agreed to the published version of the manuscript.

6.2. Funding

6.3. Acknowledgements

6.4. Ethical Approval

Ethical approval was obtained from the Ministry of Health and Tokat Gaziosmanpaşa University, Turkey, and the protocol number was 83116987-278.

6.5. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.6. Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7. References

[1] Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. Nature Medicine, 26(4), 450–452. doi:10.1038/s41591-020-0820-9.

[2] Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., … Wyles, M. D. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell, 182(4), 812–827.e19. doi:10.1016/j.cell.2020.06.043.

[3] Volz, E., Hill, V., McCrone, J. T., Price, A., Jorgensen, D., O'Toole, Á., … Nascimento, F. F. (2021). Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. Cell, 184(1), 64–75.e11. doi:10.1016/j.cell.2020.11.020.

[4] A Rambaut, N Loman, O Pybus, W Barclay, J Barrett, A Carabelli, T Connor, T Peacock, DL Robertson, E Volz, on behalf of COVID- Genomics Consortium UK (CoG-UK) (2020), Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations, Available online: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (accessed on February 2021).

[5] Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., … de Oliveira, T. (2020). Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. doi:10.1101/2020.12.21.20248640.

[6] Galloway, S. E., Paul, P., MacCannell, D. R., Johansson, M. A., Brooks, J. T., MacNeil, A., … Dugan, V. G. (2021). Emergence of SARS-CoV-2 B.1.1.7 Lineage — United States, December 29, 2020–January 12, 2021. MMWR. Morbidity and Mortality Weekly Report, 70(3), 95–99. doi:10.15585/mmwr.mm7003e2.

[7] Annavajhala, M. K., Mohri, H., Wang, P., Zucker, J. E., Sheng, Z., Gomez-Simmonds, A., … Uhlemann, A.-C. (2021). A Novel and Expanding SARS-CoV-2 Variant, B.1.526, Identified in New York. doi:10.1101/2021.02.23.21252259.

[8] Guo, W., Li, M., Dong, Y., Zhou, H., Zhang, Z., Tian, C., … Hu, D. (2020). Diabetes is a risk factor for the progression and prognosis of COVID-19. Diabetes/Metabolism Research and Reviews, 36(7). doi:10.1002/dmrr.3319.

[9] Richardson, S., Hirsch, J. S., Narasimhan, M., Crawford, J. M., McGinn, T., … Davidson, K. W. (2020). Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA, 323(20), 2052. doi:10.1001/jama.2020.6775.

[10] Hajifathalian, K., Kumar, S., Newberry, C., Shah, S., Fortune, B., Krisko, T., … Sharaiha, R. Z. (2020). Obesity is Associated with Worse Outcomes in COVID‐19: Analysis of Early Data from New York City. Obesity, 28(9), 1606-1612. doi:10.1002/oby.22923

[11] P Allison (2014), Event History Analysis: Regression for Longitudinal Event Data, 2nd edition (SAGE Publications).

[12] Fine, J. P., & Gray, R. J. (1999). A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association, 94(446), 496–509. doi:10.1080/01621459.1999.10474144.

[13] Sutradhar, R., & Austin, P. C. (2018). Relative rates not relative risks: addressing a widespread misinterpretation of hazard ratios. Annals of Epidemiology, 28(1), 54–57. doi:10.1016/j.annepidem.2017.10.014.

[14] Oulhaj, A., Ahmed, L. A., Prattes, J., Suliman, A., Alsuwaidi, A. R., Al-Rifai, R. H., … Van Keilegom, I. (2020). The competing risk between in-hospital mortality and recovery: A pitfall in COVID-19 survival analysis research. doi:10.1101/2020.07.11.20151472.

[15] Zuccaro, V., Celsa, C., Sambo, M., Battaglia, S., Sacchi, P., Biscarini, S., … Bruno, R. (2021). Competing-risk analysis of coronavirus disease 2019 in-hospital mortality in a Northern Italian centre from SMAtteo COvid19 REgistry (SMACORE). Scientific Reports, 11(1). doi:10.1038/s41598-020-80679-2.

[16] Lau, B., Cole, S. R., & Gange, S. J. (2009). Competing Risk Regression Models for Epidemiologic Data. American Journal of Epidemiology, 170(2), 244–256. doi:10.1093/aje/kwp107.

[17] Latouche, A., Allignol, A., Beyersmann, J., Labopin, M., & Fine, J. P. (2013). A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. Journal of Clinical Epidemiology, 66(6), 648–653. doi:10.1016/j.jclinepi.2012.09.017.

[18] Austin, P. C., Lee, D. S., & Fine, J. P. (2016). Introduction to the Analysis of Survival Data in the Presence of Competing Risks. Circulation, 133(6), 601–609. doi:10.1161/circulationaha.115.017719.

[19] Putter, H., Schumacher, M., & van Houwelingen, H. C. (2020). On the relation between the cause‑specific hazard and the subdistribution rate for competing risks data: The Fine‑Gray model revisited. Biometrical Journal, 62(3), 790-807. doi:10.1002/bimj.201800274.

[20] Ioannidis, J. P. A. (2018). The Proposal to Lower P Value Thresholds to .005. JAMA, 319(14), 1429. doi:10.1001/jama.2018.1536

[21] Colquhoun, D. (2017). The reproducibility of research and the misinterpretation of p-values. Royal Society Open Science, 4(12), 171085. doi:10.1098/rsos.171085.

[22] Levy, T. J., Richardson, S., Coppa, K., Barnaby, D. P., McGinn, T., Becker, L. B., Davidson, K. W., Hirsch, J. S., Zanos, T. P., Cohen, S. L., Debnath, S., Dominello, A. J., Falzon, L., Gitman, M., Kim, E.-J., Makhnevich, A., Mogavero, J. N., Molmenti, E. P., … Paradis, M. d. (2020). A predictive model to estimate survival of hospitalized COVID-19 patients from admission data. Cold Spring Harbor Laboratory. doi:10.1101/2020.04.22.20075416.

[23] Liang, W., Guan, W., Li, C., Li, Y., Liang, H., Zhao, Y., Liu, X., Sang, L., Chen, R., Tang, C., Wang, T., Wang, W., He, Q., Chen, Z., Wong, S.-S., Zanin, M., Liu, J., Xu, X., Huang, J., … He, J. (2020). Clinical characteristics and outcomes of hospitalised patients with COVID-19 treated in Hubei (epicentre) and outside Hubei (non-epicentre): a nationwide analysis of China. European Respiratory Journal, 55(6), 2000562. doi:10.1183/13993003.00562-2020.

[24] Li, X., Xu, S., Yu, M., Wang, K., Tao, Y., Zhou, Y., … Zhao, J. (2020). Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. Journal of Allergy and Clinical Immunology, 146(1), 110–118. doi:10.1016/j.jaci.2020.04.006.

[25] Grasselli, G., Greco, M., Zanella, A., Albano, G., Antonelli, M., Bellani, G., Bonanomi, E., Cabrini, L., Carlesso, E., Castelli, G., Cattaneo, S., Cereda, D., Colombo, S., Coluccello, A., Crescini, G., Forastieri Molinari, A., Foti, G., Fumagalli, R., … Iotti, G. A. (2020). Risk Factors Associated With Mortality Among Patients With COVID-19 in Intensive Care Units in Lombardy, Italy. JAMA Internal Medicine, 180(10), 1345-1355. doi:10.1001/jamainternmed.2020.3539.

[26] Williamson, E. J., Walker, A. J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C. E., Curtis, H. J., Mehrkar, A., Evans, D., Inglesby, P., Cockburn, J., McDonald, H. I., MacKenna, B., Tomlinson, L., Douglas, I. J., Rentsch, C. T., Mathur, R., Wong, A. Y. S., Grieve, R., … Goldacre, B. (2020). Factors associated with COVID-19-related death using OpenSAFELY. Nature, 584(7821), 430–436. doi:10.1038/s41586-020-2521-4.

[27] Yehia, B. R., Winegar, A., Fogel, R., Fakih, M., Ottenbacher, A., Jesser, C., Bufalino, A., Huang, R.-H., & Cacchione, J. (2020). Association of Race With Mortality Among Patients Hospitalized With Coronavirus Disease 2019 (COVID-19) at 92 US Hospitals. JAMA Network Open, 3(8), e2018039. doi:10.1001/jamanetworkopen.2020.18039.

[28] Salinas-Escudero, G., Carrillo-Vega, M. F., Granados-García, V., Martínez-Valverde, S., Toledano-Toledano, F., & Garduño-Espinosa, J. (2020). A survival analysis of COVID-19 in the Mexican population. BMC Public Health, 20(1). doi:10.1186/s12889-020-09721-2.

[29] Labenz, C., Kremer, W. M., Schattenberg, J. M., Wörns, M.-A., Toenges, G., Weinmann, A., Galle, P. R., & Sprinzl, M. F. (2020). Clinical Frailty Scale for risk stratification in patients with SARS-CoV-2 infection. Journal of Investigative Medicine, 68(6), 1199–1202. doi:10.1136/jim-2020-001410.

[30] Vekaria, B., Overton, C., Wisniowski, A., Ahmad, S., Aparicio-Castro, A., Curran-Sebastian, J., … Elliot, M. (2020). Hospital Length of Stay for COVID-19 Patients: Data-Driven Methods for Forward Planning. doi:10.21203/rs.3.rs-56855/v1.

[31] Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. Journal of the Royal Statistical Society: Series B (Methodological), 11(1), 15-44. doi:10.1111/j.2517-6161.1949.tb00020.x.

[32] Peng, Y., & Dear, K. B. (2000). A nonparametric mixture model for cure rate estimation. Biometrics, 56(1), 237-243. doi:10.1111/j.0006-341X.2000.00237.x.

[33] Cai, C., Zou, Y., Peng, Y., & Zhang, J. (2012). smcure: An R-package for estimating semiparametric mixture cure models. Computer Methods and Programs in Biomedicine, 108(3), 1255–1260. https://doi.org/10.1016/j.cmpb.2012.08.013.