Orjinal Araştırma / Original Article

A Study On Factors Impacting Length of Hospital Stay of COVID-19 Inpatient

COVID-19 Tanısı ile Hastanede Yatan Hastalarda Hastanede Kalış Süresini Etkileyen Faktörler Üzerine Bir Araştırma

Şirin Çetin[1], Ayse Ulgen[2], Hakan Şıvgın[3], Wentian Li[4]

[1]Tokat Gaziosmanpasa University Faculty of Medicine, Department of Biostatistics, Tokat, Turkey
[2]Girne American University Faculty of Medicine, Department of Biostatistics, Karmi, Cyprus
[3]Tokat State Hospital, Department of Internal Medicine, Tokat, Turkey
[4]The Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, USA

Abstract

Aim: Knowing the typical length of hospital stay of COVID-19 patientsand which factors affecting the stay time are important for hospital management.A factor can impact the stay time in conflicting ways: a condition may,by the same mechanism, increase the stay time for a severe patient while decrease the time for an even more severe and deceased patient.

Material and Method: 3184 COVID-19 patients from the Tokat State Hospital were examined on arrival to determine if they are required further treatment in hospital or not. Among them,404 are inpatients whose hospital stay time is available. By a simple and conditional linear regressions, we examine which factors statistically impact the hospital stay time,both overall and after taking into account the patient's age or eventual survival status.We also use cause-specific Cox proportional-hazard regression for competing risk to examine contribution from a factor to the rate of event of mortality, and to the event of discharge.

Results: Surviving ICU patients have longer hospital stay time than non-surviving ICU patients, which in turn longer than non-ICU patients. Older age within acertain range is always correlated with a longer hospital stay. The factors with strongest signal in our dataset are C-reactive protein (CRP), hemoglobin (HGB)and calcium level: increased CRP level, decreased HGB level and calcium level are associated with longer hospital stay, independent from the contribution from surviving status. Other potentially stay-time-impacting factors include d-dimer, urea, glucose, white blood cell (WBC) count, neutrophil, but these signals may not be robust against log-transformation of the stay time, or confounding with age. We also observed that glucose is more important than HbA1C or diabetes status in its influence on hospital stay time.

Conclusion: Almost all factors we collected contribute to a faster/slower mortality or discharge rate, in particular C-reactive protein and hemoglobin. Measurement of the associated factors in COVID-19 patients could be used for a better hospital bed management.

Keywords: COVID-19, Cox proportional-hazard regressions, demographic and clinical factors

Öz

Amaç: COVID-19 hastalarının hastanede kalış süreleri ve kalış sürelerini etkileyen faktörlerin belirlenmesi hastane yönetimi için önem arz etmektedir. Bir faktör, hastanede kalma süresini çelişkili şekillerde etkileyebilir: bir faktör, aynı mekanizma ile ağır bir hasta için kalış süresini arttırırken, daha ağır bir hasta için zamanı azaltabilir.

Gereç ve Yöntem: Tokat Devlet Hastanesin'e başvuran 3184 COVID-19 hastası hastaneye gelişlerinde muayene edilip ayakta veya yatarak tedavi edilmelerine göre kategorize edildiler. Bunlardan 404'ü hastanede kalış süresi belirtilmiş olan yatarak tedavi gören hastalardır. Basit ve koşullu doğrusal regresyonlar kullanılarak, hem genel olarak ve hem de hastanın yaşı veya hayatta kalma süresi bir faktörün ölüm oranına ve taburcu olma durumuna katkısı dikkate alınarak, hastanede kalış sürelerini etkileyen faktörler incelendi. Ayrıca, yarışan riskler için Cox orantılı hazard modeli kullanılarak bir faktörün ölüm oranına ve taburcu olma durumuna katkısı incelendi.

Bulgular: Hayatta kalan Yoğun Bakım Ünitesi hastalarının, hayatta kalmayan Yoğun Bakım Ünitesi hastalarına ve Yoğun Bakım Ünitesi olmayan hastalara göre daha uzun hastanede kalış süresine sahip olduğu bulgulandı. Artan yaşın, hastanede daha uzun kalış süresiyle ilişkili olduğu gözlemlendi. Yüksek C-reaktif protein (CRP), düşük hemoglobin (HGB) ve kalsiyum seviyelerinin, hayatta kalma durumunun katkısından bağımsız olarak, hastanede daha uzun kalma süresiyle ilişkili olduğu belirlendi. Hastanede kalış süresini etkileyen diğer faktörler arasında d-dimer, üre, glikoz, beyaz kan hücresi (WBC) sayısı ve nötrofil de bulunmaktadır ancak bu faktörler kalış süresinin log transformasyonu veya yaşla ilgili karıştırıcı etkilere karşı stabil olmayabilirler. Ayrıca, glukozun, hastanede kalış süresi ile ilgili olarak, HbA1C veya diyabet durumundan daha önemli olduğu gözlemlendi.

Sonuç: Çalışmamızda hemen hemen bütün faktörlerin ve özellikle C-reaktif protein ve hemoglobinin daha hızlı/daha yavaş ölüm veya taburcu olma oranına katkıda bulunduğu gözlemlenmiştir. İlgili faktörlerin COVID-19 hastalarında daha iyi ölçülmesi, daha iyi bir hastane yatak yönetimi için kullanılabileceğini düşünmekteyiz.

Anahtar Kelimeler: COVID-19, Cox orantılı hazard modeli, demografik ve klinik faktörler

INTRODUCTION

The COVID-19 pandemic since the year of 2019 infected 126 million people worldwide and caused 2.77 million deaths so far (March 26, 2021 data from www.worldometers.info/coronavirus/). This modern pandemic is often compared to the 1918 pandemic as both being a once in every century event in term of its global scale and, to a lesser extent, the number of deaths caused.[1] One particular issue in the COVID-19 pandemic is that, due to the fast transmission of virus by asymptomatic carriers, the number of patients can increase very fast in a short period of time. Even if only a percentage of them need medical care, the hospitals can be flooded with patients requiring beds, and an even smaller percentage needs Intensive Care Unit (ICU) and ventilators. The number of both the regular beds and ICU beds in a hospital are limited, and during peak time for the number of infected, a hospital can be in a crisis. To manage the crisis, it is very important to know what patient's information can potentially be used to predict their hospital stay length. With that information, it is possible to anticipate if the number of beds are sufficient.

A hospitalized patient may have two outcomes: mortality/death and discharge. Although there is a third possibility at the time of data collection, i.e., the patient is still in the hospital, it is less common for COVID-19 patients. A COVID-19 patient may be either discharged or die in the matter of days. If the two outcomes are treated separately, the time-to-event data is usually dealt with by the competing-cause or competing-risk survival analysis.[2] For example, one can first focus on the time-to-mortality data, and consider the discharged- from-hospital as right-censored data.[2-5]

For the issue of hospital stay, the two types of events are equivalent: a bed will become available whether the patient is discharged or has died. Therefore, the competing-cause survival analysis can become a regular survival analysis if there are right-censored data. If there are no right-censored data, the analysis strategy is even simpler: it is a simple analysis relating a factor, which can either be categorical/discrete or continuous, and the hospital stay time. We have carried out both linear regression and cause-specific survival analysis of competing risks, for a Turkish cohort with more than 400 inpatients. The advantage of the first analysis is that it models the hospital stay time more directly from a factor, though it does not give a clear explanation of why. The advantage of the second analysis is that it clearly distinguishes the two types of event (mortality and discharge), both related to hospital stay time, and provides a clear argument on why a factor may affect the two different events. The goal of our work is to determine factors influencing the hospital stay time.

MATERIAL AND METHODS

Patient data: Initially we have n=3184 COVID-19 patient information from the Tokat State Hospital. In all COVID-19 outpatients or inpatients, initially examinations were made on arrival to the hospital. As a result of these examinations in the hospital, a patient was decided to be hospitalized and treated as inpatients, or as an outpatient.

Excluding outpatients, we have n=404 inpatients whose hospital stay time is available. The patients were treated in COVID-19 regular wards with COVID-19 services, intermediate care, and intensive care units (ICU), run by different departments. All patients are confirmed to be infected with SARS-CoV-2 virus by reverse-transcriptase-polymerase-chain-reaction (RT- PCR) test of nasopharyngeal swab. Patients are excluded from this dataset due to one of the following reasons: (1) those who left the hospital against doctor's advice; (2) those who were transferred to other hospitals; (3) those who remained in hospital on April 1, 2021. The time to event (either mortality or discharge from hospital) is measured in the unit of days. The use of the data is approved by the ethics committee of the Tokat Gaziosmanpa¸sa University, as well as by the Turkey Ministry of Health (01.04.2021; 83116987/376).

Admission data included demographics (gender, age), diabetes status, blood test results (d-dimer, Alanine Aminotransferase (ALT), Aspartate Aminotransferase (AST), Lactate De-hydrogenase (LDH), urea, C-reactive protein (CRP), hemoblobin A1C (HbA1C), calcium, chlo- rine, potassium, sodium, creatinine (CR), hemoglobin (HGB), white blood cell count (WBC), neutrophil cell count (NEU), lymphocyte cell count (LYM), platelet count (PLT). Blood test measurements were done within the first 24 hours of hospitalization. We have also added these derived quantities: NEU/LYM ratio (NLR), BUN/CRE ratio (BCR) (BUN for blood urea nitrogen), and PLT/LYM ratio (PLR).

Linear regression: Denotes y for a patient's hospital stay (in days), and x is a factor. The factor can be binary/discrete/categorical, then it is coded as 1 for presence and 0 for absence; and the factor can also be continuous. The impact of a factor on the length of hospitalization can be studied by the linear regression: y=a+bx. The sign of b provides information on whether the presence of a factor (for binary factor) or an increase of the factor value (for continuous factor) lengthens the hospital stay or shortens it. The p-value concerns the test of null hypothesis b=0. Although traditionally, p-value is often judged to be significant if it is smaller than 0.01 or 0.05, we adopt a more stringent requirement for p-value<0.005. The reasons for this requirement can be found in (e.g.).[6-8] If a continuous factor x does not follow a normal distribution, whereas its log-transformed value does, we will use log(x) in the linear regression. We run linear regression twice, once with y as stay time, and another with log-stay-time.

For linear regression conditional on age, we use $y=a+bx+cx_2$ where x_2 is age. The p-value concerns the test of the null hypothesis of b=0. Similarly, For linear regression conditional on the mortality/discharge status, we use the $x_2=1, 0$ variable for these two status. The analysis was carried out by R (www.r-project.org): e.g., lm() for linear regression.

Cause-specific Cox proportional hazard regression for competing-risk survival analysis: The Cox proportional hazard model (Cox regression) assumes hazard (rate of event to occur conditional on being alive), after the time-dependent functional is removed by a ratio, and after a log-transformation, can be modeled as a

linear regression on independent variables. The right-censored samples can be handled in the Cox regression. For cause-specific Cox regression with two types of events, one may run two rounds of regular Cox regressions. In the first round, the second type of event is treated as right-censored; and in the second round, the first type of event is treated as right-censored.[2-5,9-11] We chose cause-specific survival analysis over the subdistribution survival analysis because our two types of events are mutually exclusive.[4,5,12] The R function coxph in the R package survival are used for the Cox regressions.

Dimension reduction by UMAP: A high-dimensional space can be projected to a low- dimensional one by a dimension reduction technique. There are many techniques available and we use only one of them in this paper: uniform manifold approximation and projection, or UMAP.[13] UMAP constructs a graphic representation of a high-dimensional space, then optimizes a low-dimensional representative graph which is as similar to the original graph as possible. The R package *umap* is used for the umap run.

RESULTS

Graphic illustration of distribution of the hospital stay time: The n=404 hospitalized patients can be grouped into three types: (1) non-ICU patients; (2) ICU patients who survived; (3) ICU patients who did not survive. We use multiple graphic displays to illustrate the distribution of hospital stay time.

Figure 1A shows a simple histogram for log-transformed stay time. Though the exact shape of the histogram may depend on the bin size, it is clear that surviving ICU patients have the longest stay time, followed by the non-surviving ICU patients, and the shortest stay time belong to non-ICU patients. It should be noted that the histogram of the original stay time, without a log-transformation, does not show a normal-like distribution. An implication is that one should consider log(stay-time) in analysis where the normal assumption is used.

Figure 1B shows the cumulative distribution of hospital stay time in three groups (x-axis in log scale). Cumulative distribution always provides a smoother version of the density distribution. Our previous conclusion, that surviving ICU patients have longer hospital stay time than non-surviving ICU patients, and non-surviving ICU patients have longer stay time than non-ICU patients, can be seen clearly.

To investigate the impact of age and gender on stay time, **Figure 1C-D** show the stay time (or log-transformed stay time, in **Figure 1D**) as a function of patient's age. Three groups of samples are in three different colors: non-ICU in green, surviving ICU in blue, and non-surviving ICU in black. We have the following observations: (1) stay time increases within certain age range (e.g. 50-80); (2) the conclusion from **Figure 1A-B** is confirmed in **Figure 1D,** where the median±MAD or mean±standard deviation bars, of the three groups, show the order of (surviving ICU/blue)>(non-surviving ICU/black)>(non-ICU/green); (3) Non-surviving ICU patients have a larger variance/spread of stay time compared to other groups.

Factors impacting hospital stay time from six linear regressions: Table 1 shows the linear regression of hospital stay time or its log-transform (regardless the ending event is discharge or mortality) over individual factor values. These factor values are log-transformed because the latter is more normally distributed: ALT, AST, urea, CRP, creatinine, NEU/LYM ratio, BUN/CR ratio, and PLT/LYM ratio. The following factors are significant at 0.005 level for either stay time or log(time) or both: gender (male patients stay longer), age (older patients stay longer), ending event status (non-surviving patients stay longer), d-dimer, glucose, urea, CRP, WBC, NEU/LYM ratio, BUN/CR ratio (for all above factors, higher values lead to longer stay), calcium, HGB (for these two factors, lower values lead to longer stay).

The conclusion that non-surviving patients stay in hospital longer then surviving ones might seem to be contradictory to our previous results in **Figure 1**, where surviving ICU patients stay longer than deceased ICU patients. In fact, there is no contradiction: if non-ICU patients are included in the surviving patient group, as a whole, they stay in hospital for shorter time.

Because age and surviving status are such important factors impacting hospital stay time, we would like to check if a factor is influencing hospitalization time through age, or through surviving status. Towards this, we run the linear regression conditional on age or on event ending status. The resulting p-values are shown in **Table 1**. The factors CRP and HGB remain to be significant at 0.005 level, indicating that even after considering the age factor, or considering the surviving status, they still have great influence on the stay time.

There are other factors that are significant at 0.01 or 0.05 levels, either in univariate regression or in conditional regression in **Table 1**, e.g., ALT, neutrophil, lymphocyte, platelet, platelet/lymphocyte ratio. These results may need a larger sample size data to confirm, or need a more careful analysis conditional on other factors. Note that the purpose of carrying out multiple similar regressions/tests is to examine the robustness of a result, and not a fishing expedition to increase the chance to find significant result. They should not be treated with multiple testing corrections.

Factors influencing faster mortality or faster discharge by cause-specific Cox proportional hazard regression: Because contribution from a factor to hospital stay time is essentially rooted in its contribution to severity of the disease, we have two conflicting situations. The first is when severity exceeds a break point and the patient dies quickly. The second is when the severity is not insurmountable and patient still has a fighting chance. The end result can either result in the survival or non-survival of the patient, but the hospitalization lasts a long time. The relationship between stay time and a factor value is a complex consequence of these two conflicting paths. In order to separate the two situations, we ran a cause-specific survival analysis for competing risks with Cox proportional hazard ratio. We ran the cause-specific Cox regression twice, once with mortality as the event and discharge as being right-censored, and another time with discharge as the event and mortality as being right-censored. This practice has been described in.[3-5,9-11]
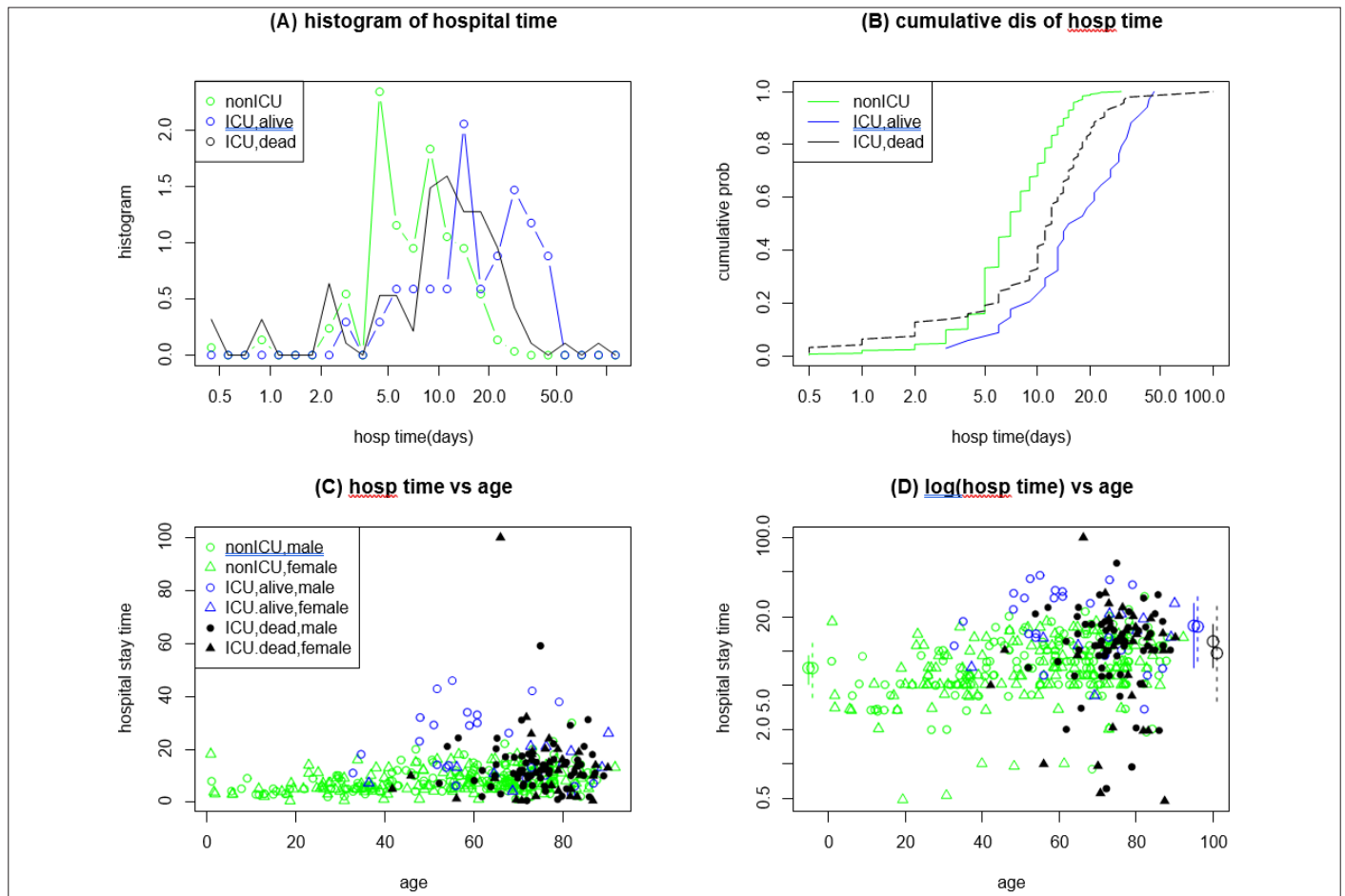
Figure 1. Various view of the hospitalization stay time in three groups: non-ICU (green), surviving ICU (blue), and deceased-ICU (black) patients: **(A)** histogram; **(B)** cumulative distribution; **(C)** stay time (y) vs. age (x); noise is added to both x and y value; **(D)** log-stay-time (y) vs. age (x). Zero stay time is converted to 0.5 and noise is added to both x and y. (ICU): Intensive Care Unit

Table 1. The N column is the number of samples whose factor value is available.

Factor	N	direction	pv	pv(log-y)	pv(x+A)	pv(log-y,x+A)	pv(x+D)	pv(log-y, x+D)
gender	404	M > F	0.018	**0.00094**	0.018	**0.00094**	0.034	**0.0015**
age	404	+	**5.4e-7**	**2.7e-9**			**0.00013**	**1.8e-8**
diabetes	226		0.66	0.99	0.66	0.99	0.99	0.82
death	404	D > S	**0.00014**	0.051	**0.00014**	0.051		
d-dimer	349	+	**0.0005**	0.0054	0.022	0.21	0.046	0.023
glucose	404	+	**0.00024**	0.0043	0.027	0.31	0.015	0.022
(log) ALT	404		0.21	0.93	0.11	0.64	0.020	0.56
(log) AST	403		0.84	0.71	0.54	0.17	0.072	0.15
(log) urea	404	+	**0.0032**	0.047	0.52	0.16	0.73	0.39
(log) CRP	289	+	**5.5e-6**	**4e-6**	0.00022	0.00043	0.00015	**1.4e-5**
HbA1C	226		0.42	0.65	0.59	0.90	0.94	0.92
calcium	403	-	**5.1e-6**	**7.2e-5**	0.016	0.22	**0.0048**	**0.00039**
chlorine	391		0.71	0.81	0.53	0.59	0.28	0.74
(log) creatinine	404		0.19	0.38	0.064	0.079	0.28	0.74
(log) LDH	386		0.011	0.15	0.25	0.76	0.68	0.66
potassium	403		0.15	0.091	0.42	0.34	0.67	0.22
sodium	404		0.42	0.089	0.97	0.41	0.49	0.30
WBC	386	+	**0.0011**	**0.0023**	0.037	0.11	0.16	0.011
NEU	349		0.0076	0.0062	0.17	0.25	0.52	0.021
LYM	348		0.48	0.34	0.017	0.37	0.038	0.65
HGB	386	-	**6.2e-10**	**2.1e-06**	**3.7e-07**	**0.00093**	**1.9e-07**	**8.7e-06**
PLT	386		0.77	0.13	0.48	0.039	0.24	0.048
(log) NEU/LYM (NLR)	348	+	0.014	**0.0024**	0.56	0.43	0.94	**0.0045**
(log) BUN/CRE (BCR)	404	+	**0.0024**	0.036	0.088	0.67	0.060	0.13
(log) PLT/LYM (PLR)	348		0.32	0.19	0.025	0.99	0.040	0.35

P-values from linear regression of (stay)   (factor) (or log-factor) or (log-stay)   (factor) (or log-factor). A p-value smaller than 0.005 is marked by boldface. The next two columns are p-values for similar linear regression with age as a co-variate. The last two columns are p-values for similar linear regression with mortality status as a co-variate. The "direction" column (for significant at 0.005 level factors only) indicates if the linear regression coefficient is positive or negative. For the discrete factors, M > F means male has longer hospital stay then female, and D > S means deceased patients have longer stay than survived patients (for the unconditional linear regression). ALT: Alanine Aminotransferase, AST: Aspartate Aminotransferase, LDH:Lactate De hydrogenase
CRP: C-reactive protein , HbA1C: hemoblobin A1C , HGB: hemoglobin, WBC: white blood cell count, NEU: neutrophil cell count, LYM :lymphocyte cell count, PLT: platelet count

Table 2 and **3** shows the p-values from the two cause-specific Cox regression for competing- risk. We put a +/- sign on the p-value to indicate the direction of the hazard ratio (HR): if HR>1, it's positive, and if HR<1 it's negative. It can be seen that almost all factors are significant (testing the null hypothesis of cause-specific HR=1) at 0.005 level. The following factors are significant at the p < 1E-9 level: age, d-dimer, urea, calcium, creatinine, sodium, WBC, NEU, LYM, HGB, NLR, BCR, PLR.

We also see that the two cause-specific HRs have different "signs". This means that a factor causing a faster mortality (HR>1 when mortality is the event) would at the same time cause a slower discharge (HR<1 when discharge is the event).[4,5]

Comparing the test results between **Table 1** (linear regression of stay time over a factor) and **Table 2** and **3** (contribution of a factor to a faster non-surviving rate or to a faster releasing

rate) is harder. The two factors which stand out in linear regression, CRP and HGB, are simply among many other significant factors in cause-specific survival analysis. Further investigation is needed to understand the exact relationship between the two different types of analyses.

Two factors significantly associated with hospital stay time: CRP and HGB: In order to examine the two most significant factors influencing the stay time (**Table 1**), CRP and HGB, we plot the raw data directly in **Figure 2**, where y-axis is the hospital stay time (in log scale) and x-axis is the CRP (in log scale) or HGB. The linear regression line (after the necessary log transformation of either y or x variable) is shown in **Figure 2**, which is positive for CRP and negative for HGB. The p-values related to the CRP association is around p<E-4~E-9, and that with HGB is around p<E-4~E-9 (**Table 1**).

Table 2. Hazard ratio/p-values of the cause-specific Cox regression competing-risk survival analysis using time-to-mortality for individual factors.

| factor	HR (95% CI)	pvD	pvD(+age)	pvD(+icu)
sex	1.02 (0.667-1.56)	0.93	0.22	0.22
age	1.05 (1.03-1.07)	**6.8e-08**	0.0094	0.01
diabete	1.51 (0.922-2.46)	0.1	0.25	0.25
d-dimer	1.54 (1.34-1.77)	**9.9e-10**	0.072	0.076
glucose	1.0032 (1.002-1.005)	**6.1e-05**	**0.00085**	**0.00089**
(log) ALT	1.36 (1.19-1.56)	**8.2e-06**	0.066	0.066
ALT	1.0004 (1.00001-1.00077)	0.042	**0.00019**	**0.00021**
(log) AST	1.42 (1.28-1.59)	**2.4e-10**	0.055	0.057
AST	1.0002 (1.00003-1.00034)	0.022	**8.1e-05**	**8.6e-05**
(log) urea	3.35 (2.62-4.27)	**3.3e-22**	0.13	0.13
urea	1.008 (1.006-1.01)	**1.7e-19**	**0.00094**	**0.001**
(log) CRP	1.1 (0.97-1.25)	0.14	0.063	0.061
CRP	1.004 (1.001-1.006)	**0.0025**	**0.0014**	**0.0013**
HbA1C	1.13 (1.03-1.25)	0.013	0.048	0.048
calcium	0.5 (0.419-0.6)	**6.2e-14**	0.098	0.097
chlorine	1.08 (1.05-1.12)	**7.9e-08**	0.015	0.014
(log) cretinine	2.62 (2.07-3.32)	**7.4e-16**	0.13	0.12
cretinine	1.372 (1.259-1.495)	**5.6e-13**	0.046	0.049
(log) LDH	1.97 (1.66-2.34)	**4.7e-15**	0.092	0.1
LDH	1.0001 (1.00002-1.0001)	0.016	**4.2e-05**	**4.5e-05**
potassium	1.41 (1.16-1.71)	**0.00056**	0.096	0.093
sodium	1.11 (1.08-1.15)	**8.1e-11**	0.017	0.016
WBC	1.08 (1.06-1.1)	**2.1e-14**	0.01	0.011
NEU	1.08 (1.06-1.1)	**2.2e-16**	0.01	0.011
LYM	0.25 (0.163-0.39)	**6.6e-10**	0.23	0.19
HGB	0.86 (0.786-0.946)	**0.0018**	0.048	0.049
PLT	0.997 (0.995-0.999)	**0.0016**	**0.00095**	**0.00083**
(log) NEU/LYM (NLR)	2.71 (2.21-3.31)	**3.5e-22**	0.11	0.11
(log) BUN/CRE (BCR)	1.84 (1.3-2.6)	**0.00056**	0.19	0.18
(log) PLT/LYM (PLR)	1.86 (1.46-2.36)	**3.5e-07**	0.12	0.11

The time-to-discharge events are treated as right-censored. The last two columns are p-values from a similar run conditional on age, or on ICU as a co-variate. P-values smaller than 0.005 are marked as boldface. For some factors, both the log-transformed value and the original level are used for the survival analysis.

Table 3. Similar for Table 2 for time-to-discharge event cause-specific survival analysis for competing risks.

factor	HR (95% CI)	pvL	pvL (+age)	pvL (+icu)
sex	1.43 (1.1-1.8)	**0.0019**	0.22	0.22
age	0.97 (0.97-0.98)	**1.4e-27**	0.0094	0.01
diabetes	0.8 (0.58-1.1)	0.17	0.25	0.25
ddimer	0.51 (0.44-0.61)	**3e-15**	0.072	0.076
glucose	0.99 (0.99-1)	**8.7e-08**	**0.00085**	**0.00089**
(log) ALT	0.89 (0.8-0.99)	0.029	0.066	0.066
ALT	0.997 (0.994-0.999)	**0.0022**	**0.00019**	**0.00021**
(log) AST	0.68 (0.58-0.8)	**2e-06**	0.055	0.057
AST	0.993 (0.989-0.997)	**0.0014**	**8.1e-05**	**8.6e-05**
(log) urea	0.46 (0.38-0.54)	**1.2e-19**	0.13	0.13
urea	0.982 (0.978-0.987)	**3.5e-14**	**0.00094**	0.001
(log) CRP	0.82 (0.77-0.87)	**1.2e-09**	0.063	0.061
CRP	0.992 (0.988-0.995)	**2.4e-06**	**0.0014**	**0.0013**
HbA1C	0.91 (0.84-0.99)	0.025	0.048	0.048
calcium	1.8 (1.6-2)	**3.1e-21**	0.098	0.097
chlorine	0.98 (0.96-1)	0.036	0.015	0.014
(log) cretinine	0.57 (0.47-0.68)	**2.9e-10**	0.13	0.12
cretinine	0.601 (0.508-0.711)	**2.6e-09**	0.046	0.049
(log) LDH	0.31 (0.24-0.4)	**8.6e-19**	0.092	0.1
LDH	0.996 (0.995-0.997)	**1e-14**	**4.2e-05**	**4.5e-05**
potassium	0.73 (0.63-0.85)	**7.3e-05**	0.096	0.093
sodium	0.95 (0.93-0.97)	**1.6e-05**	0.017	0.016
WBC	0.86 (0.83-0.89)	**1.1e-17**	0.01	0.011
NEU	0.84 (0.8-0.87)	**1.2e-17**	0.01	0.011
LYM	1.19 (1.1-1.3)	**1e-04**	0.23	0.19
HGB	1.27 (1.2-1.3)	**7.3e-19**	0.048	0.049
PLAT	1.0006 (0.9997-1.0001)	0.19	**0.00095**	**0.00083**
NEU/LYM (NLR)	0.54 (0.48-0.61)	**9.2e-24**	0.11	0.11
BUN/CRE (BCR)	0.55 (0.43-0.69)	**3.5e-07**	0.19	0.18
PLA/LYM (PLR)	0.81 (0.71-0.93)	**0.0025**	0.12	0.11

ALT: Alanine Aminotransferase, AST: Aspartate Aminotransferase, LDH:Lactate De hydrogenase CRP: C-reactive protein, HbA1C: hemoblobin A1C , HGB: hemoglobin, WBC: white blood cell count, NEU: neutrophil cell count, LYM :lymphocyte cell count, PLT: platelet count

From **Figure 2** we can see some patients who died/discharged on the same day he/she was admitted (we convert stay=0 day to 0.5 in order to take the log). Is it possible that their HGB level more reflects their near-end (near-mortality or near-discharge) status instead of a baseline pre-condition predictor? On the other hand, we also see that the value of HGB for these patients range between small to large without a trend. If we remove these short stayed patients, the negative slope in **Figure 2B** remains. We conclude that negative association between hospital stay and HGB is not an artifact of patients whose stay time is less or equal to one day.
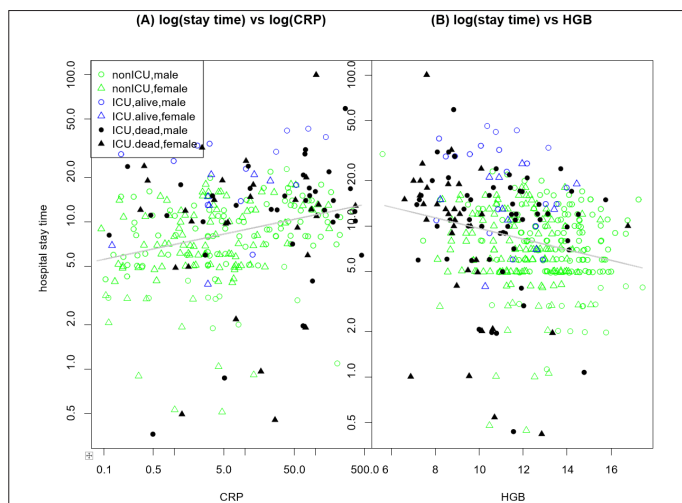
Figure 2. The hospital stay time (y) is plotted against (A) CRP (in log scale) and (B) HGB in three groups of patients: non-ICU (green), surviving ICU (blue), and deceased ICU (black). The linear regression line is also shown.

Comparing inpatients with outpatients: Although this paper is focused on hospital stay time, and outpatients would have zero stay time, we can still compare factor values between inpatients and outpatients. **Table 4** shows the mean and standard deviation of all factors (or log-transformed factor values) in four different groups: outpatients, non-ICU inpatients, surviving ICU inpatients, and deceased ICU inpatients. These factors are log transformed because they are better characterized as a single-peak normal-like distribution: ALT, AST, urea, CRP, creatinine, LDH, NLR, BUN/CR, and PLR.

To have a more complete picture of the difference between these four groups, we imagine each patient as a point in a high-dimensional space, each dimension representing a factor. That point in the high-dimensional space can be projected to a 2-dimensional space for visualization. Many dimension reduction techniques can be used, and we apply the UMAP technique.[13] After some preprocessing steps (removing factors with too much missing value, standardize each factor to have the same variance, etc.), **Figure 3** shown one version of the UMAP representation. Only these factors are used: glucose, ALT, AST, urea, calcium, chlorine,

creatinine, LDH, potassium, sodium, WBC, HGB, platelet. Only 276 outpatients are used because other had too much missing information. Those factors better represented by log- transformed version are log-transformed. **Figure 3** shows that deceased ICU inpatients tend to form a cluster which are separated from other patients.[14]
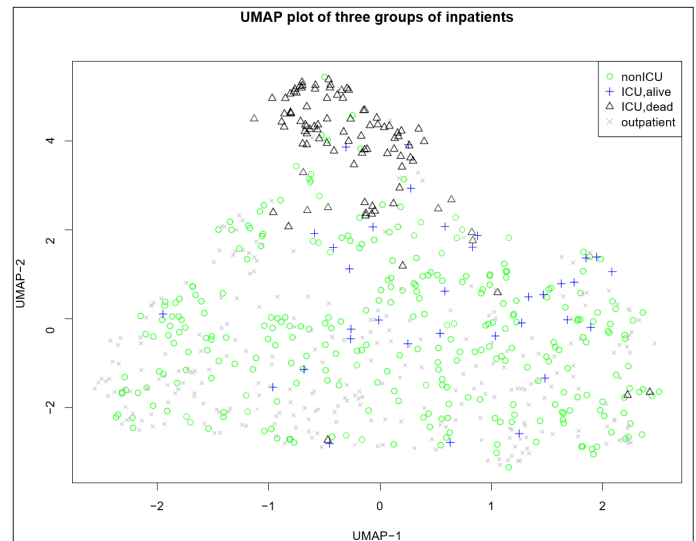
Figure 3. UMAP (McInnes et al., 2018) plot of four types of samples: outpatients (grey), non-ICU inpatients (green), surviving ICU inpatients (blue), and non-surviving ICU inpatients (black). Several preprocessing steps are taken and only a small subset of all outpatients are used. UMAP: uniform manifold approximation and projection, (ICU): Intensive Care Unit

Diabetes and stay time: Diabetic disease is shown to be a comorbidity for COVID-19, however its impact on hospital stay time is unclear.[15] We plot the stay time vs glucose and HbA1C level in **Figure 4**. It can be seen that no trend is seen in HbA1C plot, as patients with high HbA1C level spread from short (die faster) to long stay time. However, there is a positive trend for glucose level. The p-value testing the slope to be zero is significant (see **Table 1**). On the other hand, the trend in **Figure 4A** seem to be an average result, and it would be very hard to predict individual's stay time based on glucose level alone.

In a recent study, COVID-19 patients with high glucose (hyperglycemia) tend to have worse outcome, regardless of their diabetes status.[16] In other words, long-term and short-term blood sugar level should be treated separately, which is consistent with our finding in **Figure 4**. In order to further examine the connection between glucose and HbA1C, we carry out the bi-variate regression: (stay)~(glucose) +(A1C), or (log-stay)~(glucose) +(A1C). In both regressions, the p-values for the glucose term (p-value=0.03 and 0.14) are more significant than that for the HbA1C term (p-value=0.75 and 0.75). These results cast doubt on the contribution from diabetes to COVID-19 severity (and the consequent hospital stay time). On the other hand, it is not impossible that high glucose level is a consequence of COVID-19 instead of a cause, and an association analysis is not the same as causal inference.

Table 4. The average level (mean and standard deviation) for factors (or log-factors if the log-transformed variable is closer to normal distribution) of individual factors for four groups of patients: outpatients, non-ICU inpatients, surviving ICU patients, and deceased ICU patients.

Factor	outpatient mean±sd	non-ICU mean±sd	ICU-surviving mean±sd	ICU-deceased mean±sd
age	37.253±14.915	54.037±21.554	64.353±15.501	74.17±9.174
d-dimer	0.438±0.718	0.616±0.873	1.169±1.121	2.366±1.426
glucose	107.355±43.681	124.591±59.682	132.635±52.857	197.677±101.463
log(ALT)	3.033±0.638	3.152±0.836	2.912±0.893	3.769±1.464
log(AST)	3.131±0.418	3.307±0.629	3.113±0.516	4.375±1.494
log(urea)	3.269±0.363	3.415±0.496	3.561±0.495	4.716±0.643
log(CRP)	0.227±1.616	1.59±2.038	1.817±1.795	2.79±2.205
HbA1C	6.045±1.311	6.717±1.707	6.716±1.887	7.871±2.548
calcium	9.485±0.559	9.133±0.825	8.96±0.736	7.665±0.738
chlorine	9.485±0.559	9.133±0.825	8.96±0.736	7.665±0.738
log(creatinine)	-0.361±0.329	-0.251 0.508	-0.258±0.474	0.664±0.768
log(LDH)	5.307±0.34	5.507±0.453	5.587±0.31	6.398±0.736
potassium	4.29±0.407	4.316±0.502	4.238±0.51	4.809±1.201
sodium	138.982±2.836	139.391±3.495	140.062±4.072	143.989±7.022
WBC	7.137±2.743	6.848±2.864	8.761±4.952	15.824±8.197
NEU	4.703±2.409	4.654±2.786	6.942±5.059	14.463±7.999
LYM	1.712±0.948	1.648±0.995	1.455±0.98	0.79±0.773
HGB	13.493±1.843	12.658±1.952	11.221±1.81	10.294±2.166
PLT	223.019±67.189	251.301±96.611	251.324±134.888	193.169±107.113
log(NLR)	1.025±0.728	1.012±0.781	1.555±1.042	3.063±0.909
log(urea/CR)	2.869±0.307	2.905±0.406	3.058±0.441	3.292±0.484
log(PLR)	4.961±0.49	5.079±0.59	5.251±0.652	5.605±1.13

ALT: Alanine Aminotransferase, AST: Aspartate Aminotransferase, LDH:Lactate De hydrogenase
CRP: C-reactive protein,   HbA1C: hemoblobin A1C,   HGB: hemoglobin, WBC: white blood cell count,   NEU: neutrophil cell count, LYM :lymphocyte cell count, PLT: platelet count
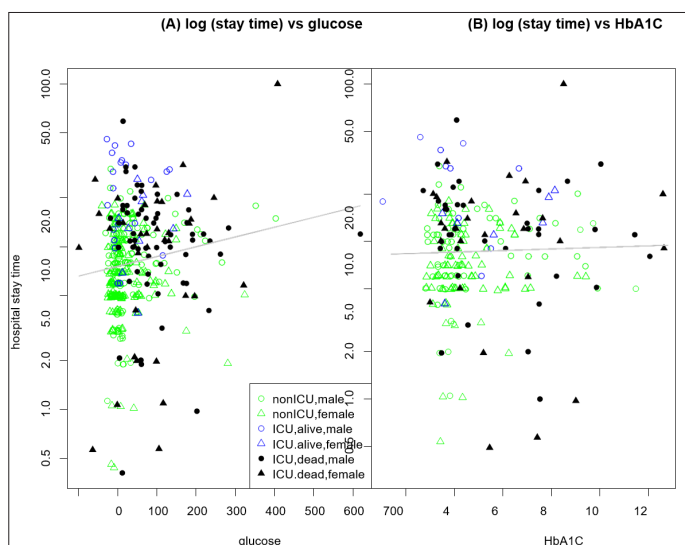
Figure 4. The hospital stay time (y) is plotted against (A) glucose and (B) HbA1C level in three groups of patients: non-ICU (green), surviving ICU (blue), and deceased ICU (black). The linear regression line is also shown.

DISCUSSION

In this paper, we aim at finding which factors are associated with long hospitalization stay time for COVID-19 inpatients. This question is related, but not identical, to the question of which factors are associated with severity of COVID-19 disease. More specifically, the associated factor should cause severity of the disease, but not too severe so that the patient dies earlier. Comparing our two associated factors, CRP and HGB, there had been reports linking CRP level to severity of COVID-19 disease, and discussion on how anemia affects the

quality of life in elder COVID-19 patients.[17-19] Interestingly, it was pointed out by a literature analysis that hemoglobin level may decrease in severe COVID-19 patients but the cause-effect direction is unclear.[20]

As for cause-specific survival analysis, we have found even more factors to be associated with the rate of mortality or rate of discharge of a COVID-19 patient. Almost all factors we measured are associated with the rate of event (**Table 2,3**). Although these results are not directly on the hospital stay time, they show potential mechanisms linking a factor with the stay time. As we have argued already, that the same mechanism can either cause a patient to stay longer in the hospital or cause another patient to stay less time. How to reconcile the results from **Table 2**, and **3** with **Table 1** worth further studies. All associations shown in **Table 2** and **3** can be found in the literature.

Besides CRP and HGB which had been discussed already, these factors are also discussed in research articles: d-dimer, glucose, urea, calcium, creatinine, potassium, sodium, WBC, neutrophile, lymphocyte, NEU/LYM ratio, BUN/CR ratio, PLT/LYM ratio.[21-34] The fact that every single one of the factors we found to be significantly associated with rate of mortality/discharge in **Tables 2** and **3** had also been discussed in the literature, not only confirm our findings, but also show an extensive dysregulation in COVID-19 patients.

Outpatients, with their zero hospitalization stay time, do not contribute to my analysis concerning stay time. However, because we have a large number of these outpatients with certain blood test results, we can still compare them with the inpatients (as in **Table 4**, and in **Figure 3** for a small proportion

of them). On average, outpatients do exhibit different blood test measurement values from either inpatients as a whole, or ICU inpatients, or deceased ICU inpatients. However, on individual person's level, we do see some outpatients being similar to non-surviving ICU inpatients (see **Figure 3**).

Our cause-specific survival analysis provides a quantitative estimation of the hospitalization time saved. Take **Table 3** for example: if HGB level increases by one unit, the time-to-discharge is on average reduced to 1/1.27 or 0.79 or the baseline population, a reduction of 21%. This is under the assumption that person eventually survives. On the other hand, if the person does not survive, increasing HGB level by one unit, the time-to-mortality is increased to 1/0.86 or 1.16 of the baseline population. By incorporating the chance for a person to survive, and considering the actual HGB level (because the impact on the rate of event may be nonlinear), it is possible to construct a calculator on the number of days saved or prolonged.

Among publications on the topic of hospital stay time for COVID-19 patients, some of them only address the overall statistics or difference between countries, on bed types and not on factors which might impact the stay time.[35] For example, it was reported that median hospital stay is 14 days in China, whereas it is 5 days outside of China.[36] Another overview paper estimates the majority of ICU hospital stay time is between 7 to 11 days.[37] These are comparable to our hospital stay values.

For studies of association between factors and hospital stay time, although the types of data varies (e.g. some only include non-severe COVID-19 patients), or the analysis plan is different (e.g., hospital stay time is binarized into longer or shorter than 14 days), we do find some interesting comparisons.[38,39] In another article, 99 surviving ("discharged") patients' hospital stay time is examined for potential association with risk factors.[40] It was observed that patients with low lymphocyte count (lymphopenia) stay in hospital longer with a log-rank test p-value of 0.027 for Kaplan-Meier curve.[40] This can be compared to our negative result concerning potential association between lymphocyte with hospital stay (**Table 1**).

Other analyses of hospital stay time applying the survival analysis techniques can be found in the literature and Factors potentially associated with hospital stay time, e.g, obesity, which may have a nonlinear impact on COVID-19 severity, the prescription of drugs ACEi and/or ARB.[41-46] Due to a lack of related information in our dataset, we are unable to compare with these published results.

CONCLUSION

Due to a potentially complex and conflicting contribution from a factor to hospitalization stay time, those factors known to be associated with COVID-19 severity cannot be automatically assumed to be associated with stay time. In this paper, we identify two factors, CRP and HGB, to be associated with stay time, and these results should be validated in other datasets.

ETHICAL DECLARATIONS

REFERENCES

1. Brown TM. The COVID-19 pandemic in historical perspective:an AJPH dossier. Am J Pub Health 2021;111:402-4.

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

3. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. Circulation 2016;133:601-9.

4. Cetin S, Ulgen A, Balci PO, et al. Survival analyses of COVID-19 patients in a Turkish cohort: comparison between using time to death and time to release. Sci Med J 2021; 3:1-9.

5. Cetin S, Ulgen A, Sivgin H, Li W, Approximate reciprocal relationship between two cause-specific hazard ratios in COVID-19 data with mutually exclusive events, medRxiv preprint, 2021; doi:10.1101/2021.04.22.21255955

6. Colquhoun D. The reproducibility of research and the misinterpretation of p-values. Royal Soc Open Sci 2017;4:171085.

7. Ioannidis JPA. The proposal to lower P value thresholds to .005. JAMA 2018;319:1429-30.

8. Li W, Shih A, Freudenberg-Hua Y, Fury W, Yang Y. Beyond standard pipeline and p<0.05 in pathway enrichment analyses. Comp Biol and Chem 2021;92:107455.

9. Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. J Clin Epid 2013;66:648-53.

10. Lau B, Cole SR, Gange SJ Competing risk regression models for epidemiologic data. Am J Epid 2009;170:244-56.

11. Putter H, Schmacher M, Van Houwelingen HC. On the relation between the causespecific hazard and the subdistribution rate for competing risks data:The Fine-Gray model revisited. Biometrical J 2020;62:790-807.

12. Fine JP and Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Am Stat Asso 1999;94:496-509.

13. McInnes L, Healy J, Saul N, Grossberger L UMAP:uniform manifold approximation and projection. J Open Source Software 2018;3:861.

14. A Ulgen,S Cetin, PO Balci, H Sivgin, M Cetin, W Li, COVID-10 outpatients and surviving inpatients exhibit comparable blood test results that are distinct from non-surviving inpatients. J Health Sci Med 2021; to appear. doi: 10.32322/jhsm.900462

15. Singh AK, Gupta R, Ghosh A, Misra A. Diabetes in COVID-19:Prevalence, pathophysiology, prognosis and practical considerations. Clinc Res Rev 2020;14:303-10.

16. Skwiersky S, Rosengarten S, Change M, et al. Sugar is not always sweet:exploring the relationhsip between hyperglycemia and COVID-19 in a predominantly African American population. J Endo Soc 2021; 5(suppl 1):A350.

17. Sharifpour M, Rangaraju S, Liu M, et al. C-Reactive protein as a prognostic indicator in hospitalized patients with COVID-19. PLoS One. 2020;15(11):e0242400.

18. Nemer DM, Wilner BR, Burkle A, et al. Clinical characteristics and outcomes of Non-ICU hospitalization for COVID-19 in a nonepicenter, centrally monitored healthcare system. J Hosp Med 2021;16:7-14.

19. Bergamaschi G, Borrelli de Andreis F, Aronico N, et al. Anemia in patients with Covid-19: pathogenesis and clinical significance. Clin Exp Med. 2021;21(2):239-246.

20. Lippi G and Mattiuzzi C. Hemoglobin value may be decreased in patients with severe coronavirus disease 2019. Hematol Transfus Cell Ther 2020;42:116-7.

21. Zhang L, Yan X, Fan Q, et al. D-dimer levels on admission to predict in-hospital mortality in patients with Covid-19. J Thromb Haemost 2020;18:1324-9.

22. Hachim MY, Hachim IY, Naeem KB, Hannawi H, Al Salmi I, Hannawi S. D-dimer, troponin, and urea level at presentation with COVID-19 can predict ICU admission:a single centered study. Front Med 2020;7:949.

23. Ceriello A. Hyperglycemia and COVID-19:What was known and what is really new? Diabetes Res Clin Pract 2020;167:108383.

24. Sun JK, Zhang WH, Zou L, et al. Serum calcium as a biomarker of clinical severity and prognosis in patients with coronavirus disease 2019. Aging 2020;12:11287-95.

25. Di Filippo L, Formenti AM, Rovere-Querini P, et al. Hypocalcemia is highly prevalent and predicts hospitalization in patients with COVID-19. Endocrine 2020;68:475- 8.

26. Benedetti C, Waldman M, Zaza G, Riella LV, Cravedi P. COVID-19 and the kidneys:an update. Front Med 2020;7:423.

27. Chen D, Li X, Song Q, et al. Assessment of hypokalemia and clinical characteristics in patients with coronavirus disease 2019 in Wenzhou, China. JAMA Netw Open 2020;3:e2011122.

28. Tzoulis P, Waung JA, Bagkeris E, et al. Dysnatremia is a predictor for morbidity and mortality in hospitalized patients with COVID-19 J Clin Endocrinol Metab. 2021;dgab107.

29. Anurag A, Jha PK, Kumar A. Differential white blood cell count in the COVID-19:A cross-sectional study of 148 patients. Diabetes Metab Syndr 2020;14:2099-2102.

30. Aschenbrenner AC, Mouktaroudi M, Krmer B, et al. German COVID-19 Omics Initiative (DeCOI). Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients. Genome Biol 2020;13:7.

31. Huang I and Pranata R. Lymphopenia in severe coronavirus disease-2019 (COVID-19):systematic review and meta-analysis. J Intensive Care 2020;8:36.

32. Liu J, Li H, Luo M, et al. Lymphopenia predicted illness severity and recovery in patients with COVID-19:a single-center, retrospective study. PLoS ONE 2020;15:e0241659.

33. Li X, Liu C, Mao Z, Xiao M, Wang L, Qi S. Predictive values of neutrophil-to-lymphocyte ratio on disease severity and mortality in COVID-19 patients:a systematic review and meta-analysis. Crtical Care 2020;24:647.

34. Simadibrata DM, Pandhita BAW, Ananta ME, Tango T. Platelet-to-lymphocyte ratio, a novel biomarker to predict the severity of COVID-19 patients:a systematic review and meta-analysis. J Intensive Care Soc 2020;to appear. doi:10.1177/1751143720969587

35. Leclerc QJ, Fuller NM, Keogh RH, et al. Importance of patient bed pathways and length of stay differences in predicting COVID-19 bed occupancy in England. medRxiv preprint 2021;doi:10.1101/2021.01.14.21249791

36. 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.

37. Lane EA, Barrett DJ, Casey M, et al. Country differences in hospitalisation, length of stay, admission to Intensive Care Units, and mortality due to SARS-CoV-2 infection at the end of the first wave in Europe:a rapid review of available literature. MedRxiv preprint 2020;doi:10.1101/2020.05.12.20099473

38. Wu S, Xue L, Legido-Quigley H, et al. Understanding factors influencing the length of hospital stay among non-severe COVID-19 patients:A retrospective cohort study in a Fangcang shelter hospital. PLoS ONE 2020;15:e0240959.

39. Hong Y, Wu X, Qu J, Gao Y, Chen H, Zhang Z. Clinical characteristics of Coronavirus Disease 2019 and development of a prediction model for prolonged hospital length of stay. Ann Transl Med 2020;8:443.

40. Liu X, Zhou H, Zhou Y, Wu X, Zhao Y, Lu Y. Risk factors associated with disease severity and length of hospital stay in COVID-19 patients. J Infection 2020;81:e95-e97.

41. Vekaria B, Overton C, Wisniowski A, et al. Hospital length of stay For COVID-19 patients:data-driven methods for forward planning. Res Square preprint 2020;doi:10.21203/rs.3.rs-56855/v1

42. Thai PQ, Toan DTT, Son DT, et al. Factors associated with the duration of hospitalisation among COVID-19 patients in Vietnam:A survival analysis. Epid & Infec 2020;148:e114.

43. Lopez-Cheda A, J´acome MA, Cao R, De Salazar PM. Estimating lengths-of-stay of hospitalized COVID-19 patients using a non-parametric model: a case study in Galicia (Spain). medRxiv preprint 2021;doi:10.1101/2020.09.04.20187963

44. Moriconi D, Masi S, Rebelos E, et al. Obesity prolongs the hospital stay in patients affected by COVID-19, and may impact on SARS-COV-2 shedding. Obesity Res Clin Pract 2020;14:205-9.

45. Kompaniyets L, Goodman AB, Belay B, et al. Body mass index and risk for COVID-19related hospitalization, intensive care unit admission, invasive mechanical ventilation, and death United States, March December 2020, Morb Mortality Wkly Rep ePub 2021;doi:10.15585/mmwr.mm7010e4

46. Braude P, Carter B, Short R, et al. The influence of ACE inhibitors and ARBs on hospital length of stay and survival in people with COVID-19. Int J Cardio Heart Vasc 2020; 31:100660.