



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Identifying and quantifying robust risk factors for mortality in critically ill patients with COVID-19 using quantile regression

Zeqiang Linli, PhD^{a,b,1}, Yinyin Chen, PhD^{c,d,e,1}, Guoliang Tian, PhD^f, Shuixia Guo, PhD^{a,b,*}, Yu Fei, PhD^{g,**}

^a MOE-LCSM, School of Mathematics and Statistics, Hunan Normal University, Changsha, PR China

^b Key Laboratory of Applied Statistics and Data Science, Hunan Normal University, College of Hunan Province, Changsha, PR China

^c Hunan Provincial People's Hospital, Hunan Normal University, Changsha, PR China

^d Changsha Clinical Research Center for Kidney Disease, Changsha, PR China

^e Hunan Clinical Research Center for Chronic Kidney Disease, Changsha, PR China

^f Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, PR China

^g School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming, PR China

ARTICLE INFO

Article history:

Received 2 July 2020

Received in revised form 31 July 2020

Accepted 26 August 2020

Keywords:

COVID-19

Mortality

Laboratory Indicator

Risk factor

Quantile regression

ABSTRACT

Objective: Many laboratory indicators form a skewed distribution with outliers in critically ill patients with COVID-19, for which robust methods are needed to precisely determine and quantify fatality risk factors.

Method: A total of 192 critically ill patients (142 were discharged and 50 died in the hospital) with COVID-19 were included in the sample. Quantile regression was used to determine discrepant laboratory indexes between survivors and non-survivors and quantile shift (QS) was used to quantify the difference. Logistic regression was then used to calculate the odds ratio (OR) and the predictive power of death for each risk indicator.

Results: After adjusting for multiple comparisons and controlling numerous confounders, quantile regression revealed that the laboratory indexes of non-survivors were significantly higher in C-reactive protein (CRP; QS = 0.835, $p < .001$), white blood cell counts (WBC; QS = 0.743, $p < .001$), glutamic oxaloacetic transaminase (AST; QS = 0.735, $p < .001$), blood glucose (BG; QS = 0.608, $p = .059$), fibrin degradation product (FDP; QS = 0.730, $p = .080$), and partial pressure of carbon dioxide (PCO₂), and lower in oxygen saturation (SO₂; QS = 0.312, $p < .001$), calcium (Ca²⁺; QS = 0.306, $p = .073$), and pH. Most of these indexes were associated with an increased fatality risk, and predictive for the probability of death. Especially, CRP is the most prominent index with an odds ratio of 205.97 and predictive accuracy of 93.2%.

Conclusion: Laboratory indexes provided reliable information on mortality in critically ill patients with COVID-19, which might help improve clinical prediction and treatment at an early stage.

© 2020 Elsevier Inc. All rights reserved.

1. Introduction

The outbreak of coronavirus-2019 (COVID-19), which the World Health Organization (WHO) labeled a “pandemic,” has spread rapidly worldwide [1]. By the end of June, COVID-19 had spread to more than 200 countries and regions, and the number of confirmed cases worldwide exceeded 10 million, with nearly 200,000 deaths, according to statistics from Johnson Hopkins University (<https://coronavirus.jhu.edu/>). Even so, the number of cases, deaths, and affected countries is still increasing. The overall death rate of patients with COVID-19 is about 5%; however, this rises to more than 25% for critically ill patients [2]. The

risk of death has usually been evaluated based on laboratory indexes [3–5]. Identification of reliable biomarkers associated with laboratory indexes may help clinicians assess risk and develop strategies for the prevention of early mortality in critical patients.

Recently, there have been many studies on the clinical course, outcomes, mortality, and risk factors of critically ill patients with COVID-19. For example, Zhou et al. [4] found that older age, higher sequential organ failure assessment score, and d-dimer greater than 1 μg/L at admission were associated with an increased probability of death. Yang et al. [6] demonstrated that the survival term of non-survivors is likely to be within 1–2 weeks after intensive care unit (ICU) admission and the older patients (>65 years) with comorbidities and ARDS are at increased risk of death. Older patients' conditions on admission such as dyspnea, lymphocytopenia, cardiovascular disease and chronic obstructive pulmonary disease (COPD), and the occurrence of acute respiratory distress syndrome (ARDS) during hospitalization were predictive of fatal outcomes [7]. Although these retrospective observations or cohort

* Corresponding author at: School of Mathematics and Statistics, Hunan Normal University, Changsha, PR China.

** Corresponding author.

E-mail addresses: guoshuixia75@163.com (S. Guo), feiyukm@aliyun.com (Y. Fei).

¹ These authors contributed equally to this work.

studies indicated the clinical courses and risk factors of severe patients via exploratory and descriptive methods, few studies have described the association between case fatality and clinical indexes while adjusting for the false discovery rates (FDR). In addition, each patient's response to the virus is different, with considerable individual differences, for example, C-reactive protein (CRP) concentrations increase with age, possibly due to subclinical conditions. Therefore, when seeking objective and general indicators, we should also consider each individual's confounding factors that are likely to influence the results of any analysis [8]. Moreover, many indexes might have a skewed distribution with outliers, for which the application of descriptive statistical methods might be limited, and robust methods are needed to precisely determine and quantify the risk factors of fatality.

Quantile regression is a robust method for addressing outliers and heteroscedasticity in response measurements, and it can estimate the conditional quantiles of the response variable to capture more information about data. Different measures of central tendency and statistical dispersion can be useful to obtain a more comprehensive analysis of the relationship between variables [9].

The current study aimed to solve two important problems, as follows: (1) identification of differences in laboratory indicators between survivors and non-survivors in critically ill patients with COVID-19 using robust statistical methods; (2) determination of risk indicators for in-hospital death and prediction of the probability of death via risk indicators. Specifically, we built a quantile regression of laboratory indexes and outcomes (survival or death) together with some other covariables to identify discrepant indicators. Then, we conducted post hoc logistic regressions to calculate the risk of fatality associated with abnormal index values. Findings from this study might help clinicians better understand this disease and reduce mortality.

2. Method

2.1. Study design and participants

This single-center study was performed at the Dabie Mountain Regional Medical Center (Huanggang City, China), which is a designated hospital for the treatment of patients with COVID-19. We included in our study all adult patients who had been diagnosed with COVID-19 according to the WHO's interim guidance, and those who were critically ill, or who died, or who were discharged between January 28, 2020 (when the first patients were admitted), and March 13, 2020. Identification of critically ill patients was achieved by reviewing and analyzing admission logs and histories from all available electronic medical records and patient care resources. The Ethics Commission of Hunan Provincial People's Hospital approved this study.

A total of 1500 COVID-19 patients were discharged from or died in the Dabie Mountain Regional Medical Center before March 13, 2020. After excluding 1272 patients who were not critically ill or who were still hospitalized, 6 patients who died within 24 h after admission, and 7 inpatients without available key information in their medical records, there were, in total 215 COVID inpatients who met our inclusion criteria (Fig. S1). Thus, those participating in this study comprised 51 patients who died during hospitalization and 164 who were discharged. Data on all patients, including demographic data, clinical symptoms, underlying diseases, and laboratory indexes were collected on admission. Laboratory indexes included information on routine tests at the time of admission, such as routine blood indexes.

2.2. Data preprocessing

Because not all patients had undergone all the laboratory examinations, we needed to preprocess the raw data. Specifically, subjects missing more than 15% clinical index values were excluded, and indexes missing more than 15% values were discarded. Accordingly, 192 patients (50 non-survivors and 142 survivors) and 30 laboratory indexes

were included in the following analysis; see Table S1 for abbreviations of the laboratory indicators. Imputation of the remaining missing values was conducted utilizing multivariate imputation by chained equations (MICE) implemented in the R package "mice" [10]. We imputed missing data for CRP (2.11%), procalcitonin (PCT; 2.11%), erythrocyte sedimentation rate (ESR; 4.74%), etc. (Table S1). Furthermore, the outliers are reset as the boundaries of the data. Specifically, the quantiles of 1% and 99% for each continuous variable were calculated, replacing the data less than 1% quantile and more than 99% quantile, respectively (see Fig. S2 for a visual representation).

2.3. Statistical analysis

The normality test and variance homogeneity test showed that laboratory indexes were highly skewed and that there was heteroscedasticity between the two groups (see Fig. 1 for a visual sense) [11,12]. It therefore would be inappropriate to use ordinary linear regression (mean regression) to identify the indicators between the two groups in this analysis. Quantile regression is a robust statistical method to capture more information about our data and work well when the data have outliers as the median is much less affected by extreme values than the mean (0.5 quantile). Therefore, quantile regression was used to identify the different laboratory indicators between the non-survival (=0) group and survival group (=1). Specifically, we built the models as follows:

$$Q_{\tau}(\text{index}) = \beta_0(\tau) + \beta_1(\tau) * \text{Group} + \beta_2(\tau) * \text{covar}_1 + \dots + \beta_{p+1}(\tau) * \text{covar}_p$$

where quantiles $\tau = 0.1, 0.3, 0.5, 0.7, \text{ and } 0.9$, covariables included age, sex, duration (time from illness onset to admission) and various baseline diseases. To better compare the beta weights for each predictor, we used the z-transformation for continuous independent and dependent variables before modeling. FDR correction was used separately for 30 indicators at each quantile (150 comparisons). In addition, a non-parametric analog of Cohen's d, quantile shift (QS) effect size, was computed for laboratory indicators that differed between the two groups to further quantify the importance of, and the difference between indicators [13]. Under normality and homoscedasticity, Cohen's $d = 0, 0.2, 0.5, \text{ and } 0.8$ correspond approximately to QS effects of 0.5, 0.55, 0.65, and 0.7, respectively [13]. For a more detailed description of the statistical methods, see Supplementary Materials. All statistical analyses in this study were implemented in R (version: 3.6.3).

2.4. Post-hoc analysis

In this section, logistic regression was used to 1) calculate the odds ratio, and 2) predict the probability of death.

Binary logistic regression was performed to quantify the risk of laboratory indicators contributing to fatality. Firstly, to avoid the effects of outliers, each laboratory index was divided into three subgroups (i.e., "Normal", "High", and "Low") according to their reference range (Table S1). Then, single-index logistic regression models were built to calculate the odds ratio (OR) for fatality. Specifically, the logistic regression was as follows:

$$\text{Logit}[P(Y = 1)] = \beta_0 + \beta_1 * \text{index} + \beta_2 * \text{age} + \beta_3 * \text{sex} + \beta_4 * \text{duration}$$

where $Y = 1$ represents the non-survival group, $Y = 0$ represents the survival group, and covariables included age, sex, and duration. The indexes included those significantly different between the two groups.

To investigate whether these indexes that differ significantly between the two groups could accurately predict probabilities (on the logit scale) of death, we used 5-fold cross-validation (CV) logistic regression to determine the prediction ability of each index. The prediction performance including accuracy, sensitivity, specificity, precision

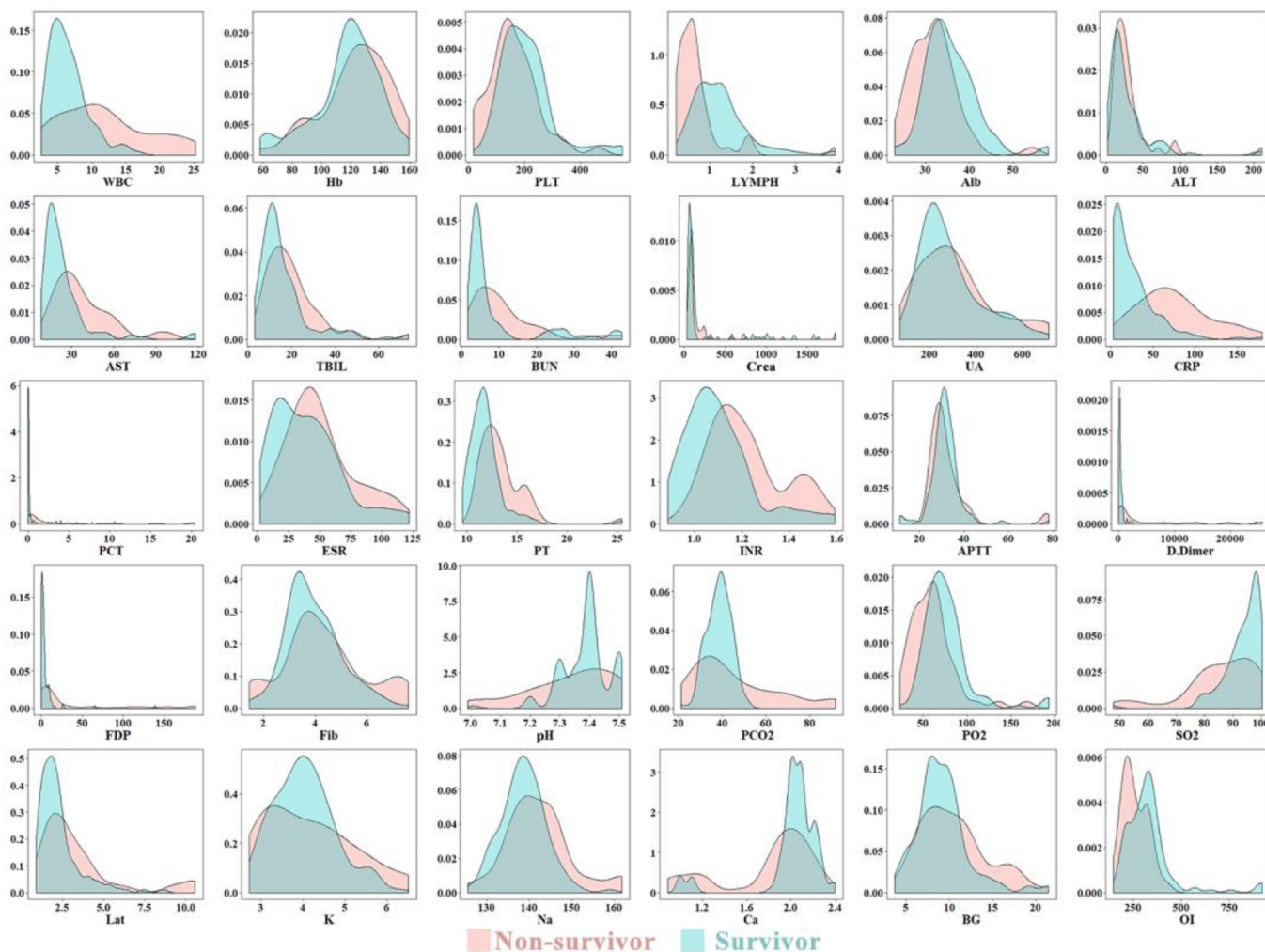


Fig. 1. Kernel density plot of 30 laboratory indexes between the two groups. Each subgraph presented a separate laboratory indicator. Non-survivors were represented in Red; survivors were represented in Blue.

rate, recall rate, and F1 score, were computed to quantify the predictive ability of the indicators. The performance measures were averaged across the 5-folds as the final performance metrics. We iterated the entire process described above ten times to avoid a biased selection of the undersampling algorithm. Detailed procedures see the Supplementary Materials.

3. Results

3.1. Descriptive statistics of self-evaluation and lab-examination indexes

Of the 192 patients with COVID-19, the mean age was 59.57 ± 16.82 , and 124 were male (64.6%). As shown in Table 1, the mean age of the non-survivor group (68.46 ± 11.74) was significantly older than the survival group (56.43 ± 17.25), and the duration between onset and admission to the hospital of non-survivors (21.08 ± 7.82) was longer than that of the other group (7.89 ± 5.33). The most common clinical symptoms were fever (155 patients, 80.7%), cough (134 patients, 69.8%), and chest pain (77 patients, 40.1%), and there is no significant difference between the two groups. However, the percentage of dyspnea in the non-survival group (17 patients, 34.0%) was significantly higher than that in the survival group (22 patients, 15.5%). Moreover, 134 patients (69.8%) had at least one underlying disease which was more obvious in the non-survival group (43 patients, 86%); non-survivors had a higher

Table 1
Demographics and baseline characteristics of critical patients with COVID-19.

	Overall (n = 192)	Non-survival (n = 50)	Survival (n = 142)	p-value
Demography				
Sex (Male)	124 (64.6%)	34 (68.0%)	90 (63.4%)	.678
Age (Mean (SD))	59.57 (16.82)	68.46 (11.74)	56.43 (17.25)	<.001
Duration (Mean (SD))	11.33 (8.39)	21.08 (7.82)	7.89 (5.33)	<.001
Symptoms -n (%)				
Fever	155 (80.7%)	42 (84.0%)	113 (79.6%)	.636
Cough	134 (69.8%)	38 (76.0%)	96 (67.6%)	.351
Myalgia	47 (24.5%)	9 (18.0%)	38 (26.8%)	.295
Diarrhea	11 (5.7%)	5 (10.0%)	6 (4.2%)	.247
Chest-pain	77 (40.1%)	26 (52.0%)	51 (35.9%)	.068
Dyspnea	39 (20.3%)	17 (34.0%)	22 (15.5%)	.010
Underlying disease -n (%)				
At least one	134 (68.8%)	43 (86.0%)	89 (62.7%)	<.001
Digestive	12 (6.2%)	7 (14.0%)	5 (3.5%)	.022
Cardiovascular	89 (46.4%)	31 (62.0%)	58 (40.8%)	.016
Cerebrovascular	34 (17.7%)	18 (36.0%)	16 (11.3%)	<.001
Hematopathy	34 (17.7%)	8 (16.0%)	26 (18.3%)	.879
COPD	36 (18.8%)	15 (30.0%)	21 (14.8%)	.031
Chronic-kidney	36 (18.8%)	7 (14.0%)	29 (20.4%)	.430
Diabetes	35 (18.2%)	6 (12.0%)	29 (20.4%)	.265

Note. Hypothesis testing was performed for comparing continuous and categorical variables in different outcome groups using independent samples t-test and Chi-squared test, respectively. Duration means time from illness onset to admission. Significant results ($p < 0.05$) were highlighted in bold.

percentage of the cardiovascular (60.2% vs. 40.8%), cerebrovascular (36.0% vs. 11.3%) and COPD (30.0% vs. 14.8%) diseases than survivors.

Regarding laboratory indexes (Table 2), descriptive statistics using the Wilcoxon test found that compared with survivors, non-survivors had significantly higher indicator values of inflammation or infection-related indicators (CRP; white blood cell counts, WBC; procalcitonin, PCT) and hypercoagulability-related indicators (D-dimer; fibrin degradation product, FDP; prothrombin time, PT). Blood gas analysis and respiratory parameters revealed more obvious hypoxia (lower partial pressure of oxygen [PO₂] and oxygen saturation [SO₂]), lactate accumulation (higher lactate [Lat]), and respiratory impairment (lower oxygenation index (OI) values) in non-survivors than in the survival group. The levels of albumin (Alb) and lymphocyte counts (LYMPH) were significantly lower, and glutamic oxaloacetic transaminase (AST) and international normalized ratio (INR) were higher in the survival group than in the non-survival group (uncorrected $p < .001$).

3.2. Differences in laboratory indices between non-survivors and survivors

Using quantile regression with $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$, we found that there were significant differences ($P_{FDR} < 0.05$) in some indicators between the two groups after adjusting sex, age, duration and underlying diseases (Figure 2(A)). Specifically, compared with survivors, individuals in the non-survival group demonstrated significantly lower indicator values in Ca²⁺ at quantile 0.5 ($\beta = -0.54$), pH at quantile 0.1 ($\beta = -1.77$) and SO₂ at quantile 0.3 ($\beta = -1.19$), while significantly higher indicator values in AST at quantile 0.5 ($\beta = 0.68$), BG at quantile 0.9 ($\beta = 1.71$), CRP at quantile 0.3–0.9 ($\beta = 1.32, 1.50, 1.73, 2.05$), FDP at quantile 0.9 ($\beta = 3.64$), PCO₂ ($\beta = 2.31$) at quantile 0.9, WBC at quantile 0.9 ($\beta = 2.19$) (Figure 2(A), Table S2). Univariate QS effect sizes of these indicators are shown in Figure 2(B) and Table S3. QS effect sizes were big and significant in AST, CRP, SO₂ and WBC ($p < .05$) and trended towards significance in BG, Ca, and FDP ($p < .1$).

Table 2
Descriptive statistic in laboratory indexes between survivors and non-survivors

Index (Reference range.)	Overall (n = 192) ^a	Non-survival (n = 50)	Survival (n = 142)	p-value ^b
WBC (4–10 × 10 ⁹ /L) ^c	6.50 [4.57, 9.60]	10.82 [7.12, 15.46]	5.81 [4.42, 7.80]	<.001
Hb (120–160 g/L)	121.00 [109.00, 134.00]	124.00 [111.25, 139.50]	119.50 [109.00, 131.75]	.079
PLT (100–400 × 10 ⁹ /L)	182.00 [134.00, 240.50]	146.00 [108.00, 200.50]	193.50 [138.00, 247.00]	.001
LYMPH (0.8–4 × 10 ⁹ /L)	0.93 [0.63, 1.37]	0.57 [0.39, 0.75]	1.11 [0.80, 1.46]	<.001
Alb (35–55 g/L)	34.00 [30.90, 37.95]	31.65 [28.10, 34.08]	34.80 [32.00, 38.80]	<.001
ALT (0–50 U/L)	20.70 [13.00, 35.20]	22.05 [15.62, 31.15]	19.00 [12.00, 37.00]	.172
AST (0–50 U/L)	22.00 [16.00, 32.15]	31.00 [22.00, 48.25]	19.00 [14.00, 27.00]	<.001
TBIL (0–20 umol/L)	13.20 [9.53, 20.50]	16.95 [11.15, 23.68]	12.40 [9.00, 19.20]	.007
Bun (1.7–8.2 mmol/L)	5.04 [3.60, 10.38]	8.88 [4.93, 13.66]	4.50 [3.45, 8.31]	.001
Crea (38–120 umol/L)	77.90 [56.82, 104.88]	89.15 [73.20, 120.60]	71.10 [55.00, 97.50]	.016
UA (204–428 umol/L)	259.65 [194.25, 364.05]	284.30 [183.25, 383.50]	250.50 [195.00, 348.93]	.407
CRP (0–10 mg/L)	25.00 [7.50, 56.50]	68.75 [45.53, 97.75]	16.10 [5.15, 34.85]	<.001
PCT (0–0.1 ng/mL)	0.10 [0.05, 0.35]	0.64 [0.24, 2.45]	0.07 [0.04, 0.15]	<.001
ESR (0–25 mm/h)	39.00 [20.00, 55.00]	44.00 [30.00, 62.00]	36.50 [18.25, 54.00]	.036
PT (8.6–12 s)	11.90 [11.00, 13.00]	13.00 [12.10, 14.10]	11.60 [10.90, 12.43]	<.001
INR (0.8–1.1)	1.10 [1.02, 1.20]	1.18 [1.11, 1.28]	1.08 [1.00, 1.16]	<.001
APTT (26–42 s)	31.00 [28.20, 34.50]	29.80 [27.30, 33.90]	31.40 [28.40, 34.52]	.27
D-Dimer (0–243 ng/mL)	312.00 [162.00, 671.00]	851.00 [328.00, 2981.00]	235.50 [144.00, 465.75]	<.001
FDP (0–5 µg/mL)	2.73 [0.85, 7.75]	8.09 [3.47, 34.41]	2.10 [0.20, 3.68]	<.001
Fib (1.9–4.6 g/L)	3.90 [3.30, 4.60]	4.20 [3.40, 5.10]	3.75 [3.20, 4.60]	.1
pH (7.35–7.45)	7.40 [7.30, 7.43]	7.39 [7.21, 7.46]	7.40 [7.35, 7.42]	.103
PCO ₂ (mmHg)	39.00 [35.00, 44.00]	38.00 [31.00, 57.00]	39.00 [35.00, 43.00]	.617
PO ₂ (mmHg)	68.00 [56.75, 82.25]	56.95 [42.03, 67.38]	72.00 [60.00, 87.50]	<.001
SO ₂ (95–98%)	94.00 [88.75, 98.00]	88.30 [80.00, 95.33]	95.00 [91.00, 98.00]	<.001
Lat (0.18–3 mmol/L)	2.10 [1.50, 2.80]	2.65 [1.90, 3.88]	1.90 [1.50, 2.50]	<.001
K ⁺ (3.8–5.4 mmol/L)	4.00 [3.40, 4.50]	3.97 [3.20, 4.76]	4.00 [3.50, 4.42]	.886
Na ⁺ (135–148 mmol/L)	139.00 [136.00, 143.00]	141.25 [137.22, 146.00]	138.40 [135.20, 142.00]	.002
Ca ²⁺ (2.25–3 mmol/L)	2.03 [1.95, 2.11]	1.95 [1.76, 2.08]	2.06 [1.99, 2.13]	.001
BG (3.9–11.1 mmol/L)	9.30 [7.30, 11.11]	9.80 [7.50, 12.35]	9.10 [7.30, 10.65]	.073
OI (400–500 mmHg)	301.00 [230.00, 344.25]	245.00 [200.00, 315.00]	320.00 [257.00, 356.00]	<.001

Note. a. All indexes were represented as median [IQR] b. Hypothesis testing was performed for comparing continuous using Wilcoxon tests. c. Indexes with p-value < .001 were shown in bold.

3.3. Results of the logistic regression

Logistic regression revealed that, after controlling for sex and duration (time from onset to admission), each additional year of age at admission was associated with a 9% increase in the odds of death (OR: 1.09). After controlling for sex and age, each additional day's duration before hospital admission was associated with a 33% increase in the odds of death (OR: 1.33); the gender effect was not significant (see Figure 2(C), Table S4 for more information). Single-index logistic regression demonstrated that increased odds of in-hospital death were associated with higher AST (OR: 4.2), BG (OR: 5.11), CRP (OR: 205.97), and WBC (OR: 7.82), and with lower PCO₂ (OR: 4.68) and pH (OR: 10.35), while controlling for age, sex, and duration. Significance for these analyses was determined at a threshold of $p < .05$, as these were post hoc analyses and included those indicators determined to be significant after FDR correction. Further correction for multiple comparisons also may have a greater public health risk of a type 2 error [14].

Next, we examined whether these discrepant indexes could predict the probability of fatality using 5-fold CV logistic regression. To be more cautious and clinically significant, subjects with a probability of fatality of more than 0.3 were deemed non-survivors in the prediction procedures. The results suggested that, over the 10 under-sampling iterations, these indicators predicted probability of death with a mean accuracy of more than 81%, mean sensitivity more than 90%, and mean specificity more than 72.8% (Figure 2(D), Table S5), in which CRP had an optimal predictive ability with an accuracy of 93.2%, sensitivity of 94.2%, and specificity of 92.2%.

4. Discussion

In this retrospective study, we focused on identifying and quantifying the relationship between fatality and laboratory indicators for 192 critical patients with COVID-19, using robust statistical methods

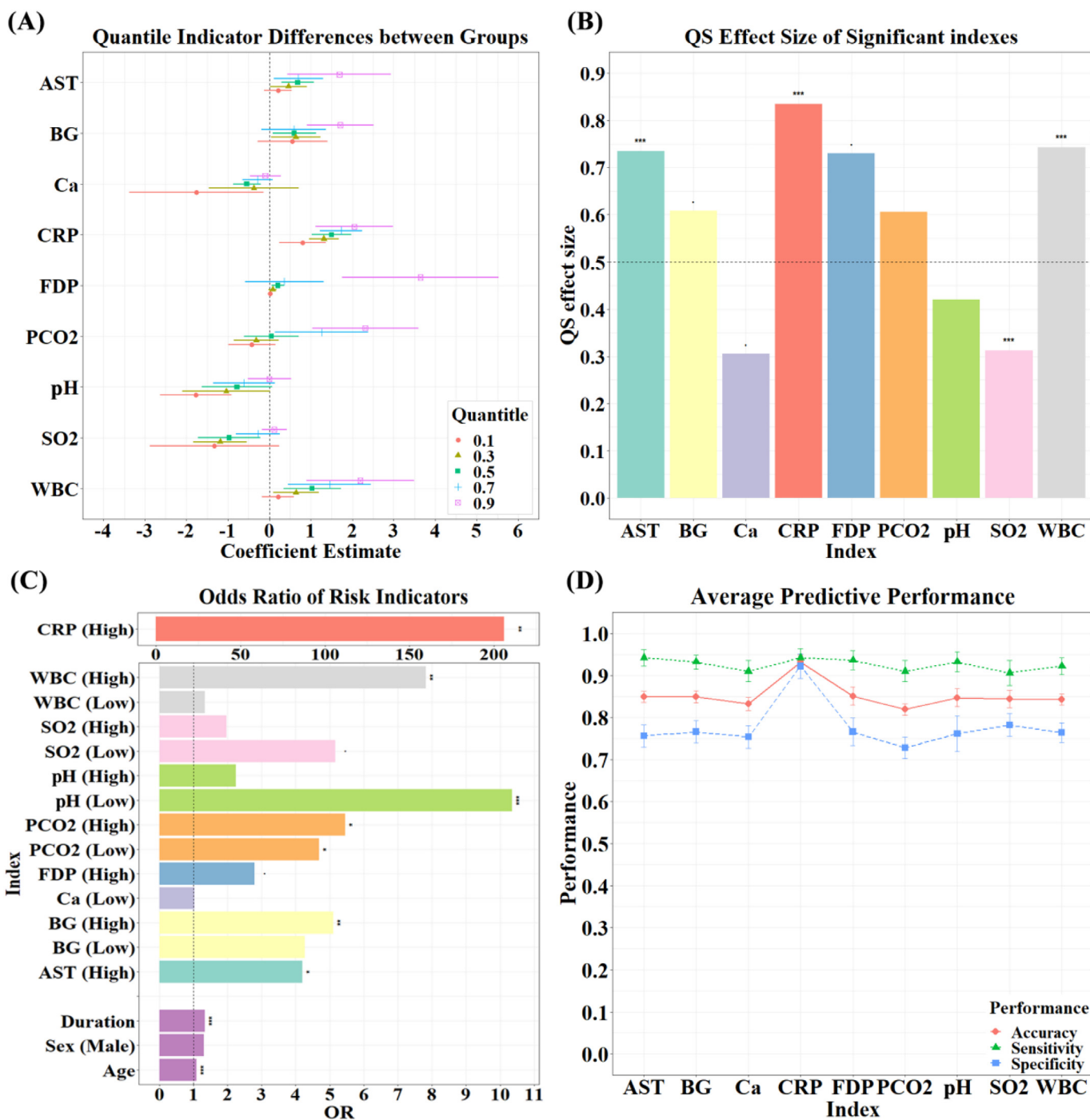


Fig. 2. (A) Beta weights of laboratory indicators that differed significantly between the two groups. (B) QS effect size of the significant indicators. Significant QS values indicate that we can reject the null hypothesis of equal distribution. QS values indicate the quantile to which the distribution has shifted from the population median (0.5). (C) The odds ratio of death by laboratory indicators: Single-index Models. “High” represented odds ratio of death by indicator values more than reference range, “Low” was similar meaning. (D) The average predictive power of indicators across 10 iterations. To be more cautious and clinically significant, subjects with a probability of fatality of more than 0.3 were deemed non-survivors in the prediction procedures.

including quantile regression and QS effect size. Several laboratory indexes were identified which provided reliable information on mortality in critically ill patients with COVID-19, which might help improve clinical prediction and treatment at an early stage.

Critically ill older patients have a higher fatality rate, as confirmed by our data [4,15]. The mean duration of non-survivors (~21 days) from onset to admission was 13 days much longer than survivors (~8 days) and each additional day was associated with a 33% increase in odds of death, indicating that early detection and early treatment are vital to

reduce the case fatality rate of critically ill patients. Although there was no significant sex difference in fatality among critical patients, recent literature has pointed out that there may be sex differences in the susceptibility and disease progression of patients with COVID-19 [16]. Therefore, we included sex as a covariate to consider more general situations. From the perspective of clinical symptoms, fever, cough, and chest pain were the most common symptoms in patients with COVID-19, consistent with the general symptoms of viral infection and pneumonia. Dyspnea was more prevalent in non-surviving patients, which

might reflect impaired respiratory function and the severity of lung lesions caused by infection or inflammation [7]. Moreover, the presence of underlying diseases can be seen among many critical patients, especially, digestive, cardiovascular, cerebrovascular and COPD diseases were more prevalent in the dead patients, which might greatly increase the vulnerability of critical patients when faced with COVID-19.

In addition to performing Wilcoxon tests for comparing laboratory variables, which discovered 14 (/30) significantly different indexes ($P_{uncorrected} < 0.001$) using comparative statistics between the dead and survivors (Table 2), we conducted quantile regressions with $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$ to more comprehensively capture data information and generate general conclusions. CRP, an important inflammation marker, was a prominent feature in our findings. CRP had consistently and significantly higher values in non-survivors than survivors among 0.3, 0.5, 0.7, 0.9 quantiles, with a considerable QS effect size of 0.835, and were significantly associated with a higher mortality risk (OR: ~206). Previous studies showed normal or slightly elevated CRP levels in the mildly ill patients [17,18] but, in our findings, most of critically ill patients (71.9%) had higher CRP levels with a median value of 25 mg/L, especially in the non-survival group with an abnormal of 98% and median values of 68.75 mg/L, which informs the severity of illness and thus prognosis. Similarly, CRP was also deemed a crucial factor contributing to the pulmonary pathology severity triggered by viral infection [19], such as H1N1 [8], H7N9 [20] and SARS [21]. Moreover, CRP was found to be the largest contributory factor among all laboratory examination indexes in another analysis that used machine learning and distinguished between the two groups (non-survivor and survivors) (revision under review). Taken together, the substantial and consistent difference in CRP between the two groups, plus the high OR values associated with death indicates that it might be a reliable biomarker of disease severity.

Increased WBC counts, AST, and reduced blood gas observations (pH, SO₂) were also noteworthy features. While normal or decreased WBC count was observed in mildly ill patients [22] or most survivors of critical illness (5.81 [4.42, 7.80]), increased WBC count was found in non-survivors with a critical illness (10.82 [7.12, 15.46]) and correlated with higher odds of death (OR: 7.82), suggesting that comorbid bacterial or fungal infection might have occurred in these deceased patients. This was consistent with our clinical observations and some reports of other viral diseases. In critically ill patients with the Middle East respiratory syndrome (MERS), 18% had bacterial coinfections [23]. Viral-bacterial coinfections in the respiratory tract increase mortality and morbidity in children and adults [24–26]. As mentioned in a previous study, nearly 100% of COVID patients who died in the ICU had sepsis [4]. The death of the non-survivors who tended to suffer from severe coinfection may be due to the need for invasive treatments such as tracheal intubation, tracheotomy or urinary tract intubation-assisted treatment. In addition, many critically ill patients had to receive a high dose of glucocorticoid treatment to suppress cytokine storm and so were then at a high risk of glucocorticoid-related secondary infections, some even developed sepsis. Previous literature has indicated acute liver damage was more frequent in COVID-19 patients than in other pneumonia patients [27] and patients with severe COVID-19 seem to have higher rates of liver dysfunction [28,29]. Huang et al. (2020) suggested that elevation of AST was observed in eight (62%) of 13 patients in the intensive care unit (ICU) compared with seven (25%) of 28 patients who did not require care in the ICU [30]. Wang et al. (2020) found that the median AST of ICU patients was 52 (30–70) U/L more than non-ICU patients with 29 (21–38) U/L [3]. These findings together with our results, suggests that liver health can be a reliable indicator of disease severity, and close monitoring and evaluation of liver function in critically ill patients should be considered. Reduced pH and SO₂ often suggest poor prognosis in pulmonary diseases, and are associated with the possibility of rapid progression to acute respiratory distress syndrome, septic shock, uncorrectable metabolic acidosis, coagulation dysfunction and even death.

This study has several limitations. First, this was a retrospective study. While our results identified and quantified risk factors for COVID-19 using robust methods, further cohort study is still needed to confirm our findings. Second, although the sample size of the analysis was close to 200, there is an unbalanced sample size only with 50 non-survivors, which limited the number of covariables used in logistic regression [31]. Third, since currently there is no clear method for evaluating multivariate effect size for quantile regressions, we only calculated univariate estimates of QS to quantify the difference [13].

In conclusion, while there have been some studies on the risk factors of COVID-19, there are few studies that have comprehensively quantified the association of case fatality and laboratory indexes while adjusting false discovery and controlling for numerous confounders. We found that abnormal values of CRP, WBC, AST, pH, SO₂, etc. associated with a higher risk of fatality, and suggest that, for those with abnormal values at these indexes, especially for CRP, close monitoring and early intervention might be very important and could help to reduce mortality.

Study approval

The Ethics Commission of Hunan Provincial People's Hospital approved this study. Registration number: ChiCTR2000030941. The Ethics Commission exempted from the need for patient consent.

Data accessibility statement

The data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

Author contribution

Drs Linli and Chen equally contributed to this work. The study was designed by Linli, Chen, Guo, and Fei. Chen collected the data, had full access to all of the data in the study and takes responsibility for the integrity of the data. Linli, Tian, Guo and Fei developed the analytical plan. Linli and Guo undertook the statistical analyses and took responsibility for its accuracy. Chen, Linli, Guo and Fei wrote the first draft of the paper. All authors contributed to writing the manuscript.

Funding

SXG is supported by the National Natural Science Foundation of China (NSFC) grant (No. 11671129, 31671134), YYC is supported by the Health and Family Planning Commission of Hunan Province, China (No. B20180428), Hunan Provincial Natural Science Foundation for Outstanding Youth (2020JJ2020), YF is supported by the National Natural Science Foundation of China (NSFC) grant (No.11971421).

Declaration of Competing Interest

All authors declare no competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ajem.2020.08.090>.

References

- [1] World Health Organization. WHO announces COVID-19 outbreak a pandemic; 11, March 2020.
- [2] Grasselli G, Zangrillo A, Zanella A, et al. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy region, Italy. JAMA. 2020. <https://doi.org/10.1001/jama.2020.5394>.

- [3] Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA*. 2020. <https://doi.org/10.1001/jama.2020.1585>.
- [4] Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3).
- [5] Chen T, Wu Di, Chen H, et al. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study. *BMJ*. 2020;368:m1091. <https://doi.org/10.1136/bmj.m1091>.
- [6] Yang X, Yu Y, Xu J, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med*. 2020. [https://doi.org/10.1016/S2213-2600\(20\)30079-5](https://doi.org/10.1016/S2213-2600(20)30079-5).
- [7] Wang L, He W, Yu X, et al. Coronavirus disease 2019 in elderly patients: characteristics and prognostic factors based on 4-week follow-up. *J Infect*. 2020. <https://doi.org/10.1016/j.jinf.2020.03.019>.
- [8] Vasileva D, Badawi A. C-reactive protein as a biomarker of severe H1N1 influenza. *Inflamm Res*. 2019;68:39–46. <https://doi.org/10.1007/s00011-018-1188-x>.
- [9] Koenker R. *Quantile regression. Econometric society monographs*. Cambridge: Cambridge University Press; 2005.
- [10] van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45. <https://doi.org/10.18637/jss.v045.i03>.
- [11] Jarque CM, Bera AK. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econ Lett*. 1980;6:255–9. [https://doi.org/10.1016/0165-1765\(80\)90024-5](https://doi.org/10.1016/0165-1765(80)90024-5).
- [12] Levene H. Robust tests for equality of variances. *Contributions to probability and statistics Essays in honor of Harold Hotelling*; 1961. p. 279–92.
- [13] Wilcoxon R. A robust nonparametric measure of effect size based on an analog of Cohen's d, plus inferences about the median of the typical difference. *J Mod Appl Stat Methods*. 2018;17:eP2726.
- [14] Salminen LE, Wilcox RR, Zhu AH, et al. Altered cortical brain structure and increased risk for disease seen decades after perinatal exposure to maternal smoking: a study of 9000 adults in the UK biobank. *Cereb Cortex*. 2019;29:5217–33. <https://doi.org/10.1093/cercor/bhz060>.
- [15] Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020;395(10223):507–13. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7).
- [16] Wei X, Xiao Y-T, Wang J, et al. Sex Differences in Severity and Mortality Among Patients with COVID-19: Evidence from Pooled Literature Analysis and Insights from Integrated Bioinformatic Analysis; 2020.
- [17] Ren L-L, Wang Y-M, Wu Z-Q, et al. Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. *Chin Med J (Engl)*. 2020. <https://doi.org/10.1097/CM9.0000000000000722>.
- [18] Zhang MQ, Wang XH, Chen YL, et al. Clinical features of 2019 novel coronavirus pneumonia in the early stage from a fever clinic in Beijing. *Zhonghua Jie He He Hu Xi Za Zhi*. 2020;43:215–8. <https://doi.org/10.3760/cma.j.issn.1001-0939.2020.03.015>.
- [19] Zheng J, Perlman S. Immune responses in influenza a virus and human coronavirus infections: an ongoing battle between the virus and host. *Curr Opin Virol*. 2018;28:43–52. <https://doi.org/10.1016/j.coviro.2017.11.002>.
- [20] Wu W, Shi D, Fang D, et al. A new perspective on C-reactive protein in H7N9 infections. *Int J Infect Dis*. 2016;44:31–6. <https://doi.org/10.1016/j.ijid.2016.01.009>.
- [21] Jang T-N, Yeh DY, Shen S-H, et al. Severe acute respiratory syndrome in Taiwan: analysis of epidemiological characteristics in 29 cases. *J Infect*. 2004;48:23–31. <https://doi.org/10.1016/j.jinf.2003.09.004>.
- [22] Xu Y-H, Dong J-H, An W-M, et al. Clinical and computed tomographic imaging features of novel coronavirus pneumonia caused by SARS-CoV-2. *J Infect*. 2020;80:394–400. <https://doi.org/10.1016/j.jinf.2020.02.017>.
- [23] Arabi YM, Mandourah Y, Al-Hameed F, et al. Corticosteroid therapy for critically ill patients with Middle East respiratory syndrome. *Am J Respir Crit Care Med*. 2018;197:757–67. <https://doi.org/10.1164/rccm.201706-1172OC>.
- [24] Meskill SD, O'Bryant SC. Respiratory virus co-infection in acute respiratory infections in children. *Curr Infect Dis Rep*. 2020;22:3. <https://doi.org/10.1007/s11908-020-0711-8>.
- [25] Kiedrowski MR, Bomberger JM. Viral-bacterial co-infections in the cystic fibrosis respiratory tract. *Front Immunol*. 2018;9:3067. <https://doi.org/10.3389/fimmu.2018.03067>.
- [26] Uyeki TM, Bernstein HH, Bradley JS, et al. Clinical practice guidelines by the Infectious Diseases Society of America: 2018 update on diagnosis, treatment, chemoprophylaxis, and institutional outbreak management of seasonal influenza. *Clin Infect Dis*. 2019;68:e1–47. <https://doi.org/10.1093/cid/ciy866>.
- [27] Zhao D, Yao F, Wang L, et al. A comparative study on the clinical features of COVID-19 pneumonia to other pneumonias. *Clin Infect Dis*. 2020. <https://doi.org/10.1093/cid/ciaa247>.
- [28] Zhang C, Shi L, Wang F-S. Liver injury in COVID-19: management and challenges. *Lancet Gastroenterol Hepatol*. 2020. [https://doi.org/10.1016/S2468-1253\(20\)30057-1](https://doi.org/10.1016/S2468-1253(20)30057-1).
- [29] Cai Q, Huang D, Ou P, et al. COVID-19 in a designated infectious diseases hospital outside Hubei Province, China. *Allergy*. 2020;75:1742–52. <https://doi.org/10.1111/all.14309>.
- [30] Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
- [31] Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49:1373–9. [https://doi.org/10.1016/S0895-4356\(96\)00236-3](https://doi.org/10.1016/S0895-4356(96)00236-3).