Decision Analytics Journal 5 (2022) 100137

Contents lists available at ScienceDirect

Decision Analytics Journal

journal homepage: www.elsevier.com/locate/dajour

A survival analysis approach for identifying the risk factors in time to recovery of COVID-19 patients using Cox proportional hazard model

Dhruba Das [a], Hemanta Saikia [b,*], Dibyajyoti Bora [c], Dibyojyoti Bhattacharjee [d], Jondeep Das [a]

[a] Department of Statistics, Dibrugarh University, India
[b] Department of Agricultural Statistics, Assam Agricultural University, India
[c] Department of Statistics, Cotton University, India
[d] Department of Statistics, Assam University, India

ARTICLE INFO

Keywords:
Coronavirus
Cox's proportional hazard model
Mantel-Haenszel test
Survival analysis
Mann-Whitney *U* test

ABSTRACT

The coronavirus pandemic was a global health crisis taking away millions of lives worldwide. People diseased by the virus, differ in the extent of severity of the infection. While it turns out to be fatal for some, for several others the extent of severity is as ordinary as common cold. These people are reported to have recovered from the disease without hospitalization and consuming some relevant medicine and home remedies. But people who have comorbidity like geriatric, high blood pressure, heart and lung problems, diabetes, cancer etc. are at high risk of developing serious illness from the infection. This study is an application of the Cox proportional hazard model with an aim to identify the risk factors that affect the recovery time of the COVID-19 patients. The model is an advanced regression technique that can be utilized to evaluate simultaneously the effect of several factors on the possibility of instantaneous failure in patients. The paper also uses the Mantel-Haenszel test (Log-Rank test) to compare if the probability of survival of different treatment procedures or different groups of patients differ significantly. The information is collected from 129 respondents of Assam, India. The study identifies that the significant risk factors that prolong the recovery time from COVID-19 are pre-disease, location, and food habits.

1. Introduction

Severe acute respiratory syndrome (SARS) coronavirus (SARS-CoV2) is a novel virus that was recognized at the end of December 2019 around the globe. It has rapidly spread from Wuhan city of China and caused a pandemic called "COVID-19" throughout the world because of its capacity to spread from human-to-human transmission. It has spread through the large droplets produced during coughing and sneezing by symptomatic individuals and may spread from asymptomatic individuals before starting the symptoms. The symptoms involved as per clinical characteristics in COVID-19 are fever, dry cough, fatigue, sore throat, nausea, vomiting, diarrhea, dyspnea, etc. Though no single features are there to distinguish COVID-19 from other viral infections, it may develop pneumonia, pulmonary failure, and death. Overall, the infection of the disease is characterized by a range of clinical presentations from asymptomatic to fatal or severe illness. The solidest forecaster for COVID-19 severity is the age over 60 years. The adults of any age also affected by this disease with underlying certain medical conditions such as cardiovascular diseases, diabetes, and chronic respiratory diseases, etc. Sometimes the patients those who immunocompromised are also at increased risk of severe illness. Moreover, the patients having comorbidity like geriatric, high blood pressure, heart and lung problems, diabetes, cancer etc. are at high risk of developing serious illness from the COVID-19 infection and the mortality rate is also higher in this segment of the patients all over world.

For any contagious disease, whether it is infectious or not, several factors are always associated with its occurrence, treatment, and recovery time. These factors usually control the recovery period and may be a cause of a prolonged recovery period for the disease. Sometimes these factors jointly affect the recovery time of the disease and hence they are termed as risk factors. The risk factors are those variables or hazards that, if it is found to be present for an individual, will develop the disease within the individual [1,2]. Owing to the risk factors, an individual might be more likely to have severe illness than others because he (or she) has physiognomies or some underlying medical conditions that increase the risk. Therefore, to recognize it as a risk factor, a variable must be associated with an increased probability of disease and must prolong the recovery period of the disease. Furthermore, the variables that may be risk factors at one life stage may or may not put an individual at risk at a later stage of development. Hence, the risk factors can reside with the individual or within the family, community, or institutions that surround the individual [3]. Sometimes these risk

* Corresponding author.
E-mail addresses: h.saikia456@gmail.com, hemanta.saikia@aau.ac.in (H. Saikia).

https://doi.org/10.1016/j.dajour.2022.100137
Received 15 June 2022; Received in revised form 1 September 2022; Accepted 13 October 2022
Available online 20 October 2022

factors may be neglected in terms of the disease due to the poor medical facility, poor treatment facilities, and so on.

The novel coronavirus was a threat to global health creating a crisis as large to be termed as a pandemic taking away millions of lives across the globe. It is mostly spread through respiratory droplets when these are inhaled or ingested by a healthy individual or transferred by hand from a contaminated surface to his/her eyes, nose, or mouth. There are also indications that people carrying the virus but without symptoms are known as asymptomatic and they can transmit the virus to other people. Though the most common symptoms of COVID-19 are fever, dry cough, and tiredness usually mild and begin gradually, some people become infected with only having very mild symptoms. These people are reported to recover from the disease without requiring hospital treatment. Also, people who have underlying medical problems like high blood pressure, heart and lung problems, diabetes, or cancer are at high risk of developing a serious illness if infected by the virus. Therefore, the study tries to identify the risk factors in the recovery time of the COVID-19 patients with mild symptoms. Also, among all the risk factors what are the factors responsible for the prolonged recovery time of the disease is examined through survival analysis. Finally, the combined effect of the significant risk factors in the recovery time of the patients is studied using *Cox's proportional hazard model*.

2. Review of literature

An individual with the existence of risk factors that are severe needs intensive care or hospitalization. Thus, identifying the risk factors for COVID-19 is indispensable because it can help one to take necessary precautions as and when they get infected by the virus. Several studies have already been performed by researchers across the world in this regard. Schleicher et al. [4] studied a case of a patient with COVID-19 infection, progressive pneumonia, development of a hyper inflammatory state, and cytokine release syndrome (CRS) who was successfully treated with steroids and tocilizumab. An observational study on Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China has been studied by Yang et al. [5]. Yuen et al. [6] have studied some most important research questions concerning virus transmission, asymptomatic and presymptomatic virus shedding, diagnosis, treatment, vaccine development, the origin of the virus, and viral pathogenesis. Hampshire et al. [7] has studied the cognitive deficits in people who have recovered from COVID-19 analyzing 81,337 individuals. A study with the principle purpose to discuss the COVID-19 disease beginning from virology, epidemiology and continuing with clinical manifestations, diagnosis, its complications and to finish with available therapeutic options has been performed by Ozdemir [8]. Norouzi et al. [9] have studied the complications related to COVID-19 disease relevant to diabetes and tried to focus on the present data and growing concepts surrounding SARS-CoV-2 infections in T2DM patients. Ugwueze [10] has also studied the link and clinical implications between the COVID-19 and diabetes mellitus. While Brosnahan [11] studied the effects of severe acute respiratory syndrome coronavirus-2 infection on different parts of the respiratory system, clues to understanding the underlying biology of respiratory disease. A study on the influence of a COVID-19 vaccine's effectiveness and safety profile on vaccination acceptance was done by Kaplan et al. [12]. Nalbandian et al. [13] performed a discussion on relevant considerations for the multidisciplinary care of COVID-19 survivors and propose a framework for the identification of those at high risk for post-acute COVID-19 and their coordinated management through dedicated COVID-19 clinics. A study on analysis of sex hormones and menstruation in COVID-19 women of child-bearing age has been done by Li et al. [14]. Chodaket et al. [15] analyzed the current state of knowledge about SARS-CoV-2 as well as its potential connection with food as a source of pathogen and infection. Mahmud et al. [16] studied the incidence, association, and risk factors associated with the development of the post-COVID-19 syndrome. A study on the effect of information about COVID-19 vaccine effectiveness and side effects on behavioral intentions was studied by Kerr et al. [17]. Augustin et al. [18] studied a longitudinal, prospective analysis of health consequences in patients who initially presented with no or minor symptoms of severe acute respiratory syndrome coronavirus type 2 (SARS-CoV2) infection.

For 844 patients from Italy with SARS-Cov-2 infection, a study by Sormani et al. [19] found that increased risk of severe COVID-19 in individuals treated with ocrelizumab or rituximab while compared to untreated people. A multi-country study was performed by Simpson-Yap et al. [20] with 2340 SARS-Cov-2 infection patients and they found that with rituximab and ocrelizumab, risk of hospitalization, ICU admission and ventilator increases. Another study based on cross-sectional survey conducted in Iran by Safavi et al. [21] for 712 patients found that enactment of anti-CD20 therapies effective for the people having fever, cough and shortness of breath. A study performed in Poland by Czarnowska et al. [22] which was relatively small and includes only 396 COVID-19 patients found that patients on ocrelizumab had a higher risk of hospitalization. An overview of factors that are possibly influencing the sharp increase of the COVID-19 outbreak in Northern Italy was performed by Goumenou et al. [23]. In case of obese adults, a high significant correlation between unhealthy eating patterns or uncontrolled eating management and immune deficiency reported by Smethers and Rolls [24]. In similar direction, Wu et al. [25] established the correlation between obesity-induced immune deficiency and COVID-19 adverse outcomes. In case of vulnerable populations, a study performed by Bornstein et al. [26] found that obesity, increases food intake, and nutrient imbalance affect immune deficiency in a bidirectional way. However, in a retrospective study with 1488 patients by O'Brien et al. [27] found that patients in US with lower body mass index were associated with higher mortality in COVID-19 as compared obese patients. Petrakis et al. [28] tried to examine the obesity as a risk factor for increased COVID-19 prevalence, severity and lethality. According to Sanyaolu et al. [29] diabetic individuals have higher morbidity, mortality and hospitalization and ICU admission. Moreover, cancer patients are more likely to get infected in COVID-19 due to immunocompromised status instigated by radiotherapy, hormonal therapy, and surgery etc. [30]. As per the study performed by Derosa et al. [31] age is more important for the patients those who infected in COVID-19 than cancer. It is because pediatric cancer patients show relative resistance to COVID-19 infection. With an analysis of 637 patients Li et al. [32] revealed that 30% of cases have a high risk of cardiovascular ailments for COVID-19 infection and 50% cases have a high prevalence of diabetes and high blood pressure. Though mild heart injury persistent cardiovascular damage is basic illness associated with COVID-19 infection but a retrospective study by Shi et al. [33] with 1009 COVID-19 patients found that 14.9% of the patients suffered from hypertension. Further Zhou et al. [34] reported with 72 314 COVID-19 patients 10.5% death rate among the people with cardiovascular damage and also found that patients with coronary artery disease as well as hypertension are more prone to SARS-CoV-2 infection. However, adolescents with COVID-19 seem to have tolerant symptoms of COVID-19 than adults as adolescents' immune systems are more capable of responding to novel germs than the adults. This capability is diminished with the advancement of age in patients with 70 years old and they even may be ineffective [35]. As per the study Chen et al. [36] the African heritage adolescents especially develop gastrointestinal indications and Kawasaki disease shock syndrome due to COVID-19.

In COVID-19 patients, cardiac injury is exposed in multiple ways. The contamination in the cardiovascular system of the patients is more vulnerable to heart attack. It has occurred when due to lack of oxygen supply to various organs of the human body [37]. During fever and inflammation for COVID-19 patients, since infection is contained in the lung initially so stress level increases and due to the characteristics of inflammatory pathway it leads to damage to the cardiac muscle and heart failure [38]. Impact of COVID-19 can also be seen in the patients who is suffering from hypertension. Lippi et al. [39] demonstrated a 2.5-fold more risk of lethality in COVID-19 with high blood pressure. Based on the data collected from several cardiological societies [40] found that use of drugs like ACE inhibitors and ARB in chronically elevated blood pressure in COVID-19 hospitalized patients. Spudich and Nath [41] proposed that younger people with mild disease could develop acute COVID-19 and long COVID neuropsychiatric syndrome. A systemic review and meta-analysis indicated that there are 2.5 times higher risk of severe infection among subjects with pre-existing stroke [42]. According to the study by Ruan et al. [43] where 179 subjects were diagnosed with SARS-Cov-2 pneumonia, prior cardiovascular difficulties significantly boosted mortality. To assess the evidence on SARS-CoV2 infection and COVID-19 in relation to deficiency and supplementation of vitamin D, D'Ecclesiis et al. [44] performed a systematic review and meta-analysis based on the data up to April 2021. A review study was conducted by Peng et al. [45] to examine the role of hypertension on the severity of COVID-19. Liu et al. [46] studied the adverse cardiovascular effects of anti-COVID-19 drugs through the review of the relative therapeutic efficacy. The crucial role of the nose on COVID-19 was discussed by Gamerra et al. [47]. A systematic review was performed by Matias et al. [48] based on metabolic syndrome and COVID-19 which interferes in greater severity and mortality through several factors.

Some of the research mentioned above is to identify the risk factors for COVID-19 disease, but these studies include children and adults with severe illness resulting in hospitalization and the Intensive Care Unit (ICU). So far, to the best of our exploration, none of the studies in the existing literature tried to examine the characteristics of individuals between symptomatic and asymptomatic patients with COVID-19. Furthermore, another gap has been identified, to study how individuals who experienced severe illness with COVID-19 are different from a mild illness. Therefore, this study has especially focused on individuals who had met mild illness and were not admitted to the ICU or the hospital.

3. Objectives of the study

On this backdrop through the literature survey, a few risk factors for COVID-19 patients are identified for the study and they are gender, age, location, vaccination status, pre-disease condition, place of treatment, and food habits. These risk factors are found to be relevant where the study is being conducted. Based on these risk factors, the study tries to examine the effect of these risk factors on the recovery time of COVID-19 patients with mild illness. Though there are a few cross-sectional studies based on COVID-19 patients but in contrast, most of the studies conducted by the researchers discussed in literature review are based on clinical study. These clinical studies have used standard medical procedures to identify the responsible factors like hypertension, heart disease, etc. However, we are using survival analysis to identify the risk factors in time to recovery of mild affected COVID-19 patients through cross-sectional survey. The use of survival analysis to identify the risk factors through cross-sectional data is the novelty as well as originality of this research. Consequently, the following three objectives are specified.

(a) To examine the case control risk factors that lead to COVID-19 infection.
(b) To identify the risk factors that may prolong the recovery time of COVID-19 patients.
(c) To study the impact of risk factors on the recovery time of COVID-19 patients.

4. Data collection

For this case control study, the relevant data have been collected from different places of Assam (a province in the north-eastern part of India). It has occupied an area of 78,438 square kilometers with a population density of 397 people/square kilometers. For data collection, a structured questionnaire is prepared and it has been validated and modified through a pilot survey from 30 patients who got affected by COVID-19 but recovered from the disease. The convenient sampling technique is used to collect the data in the pilot survey. Thereafter, the prepared structured questionnaire is being sent to 225 patients through email and finally, only 129 questionnaires were received from the mild infectious COVID-19 patients of Assam. The time-period of data collection is from 09 May 2021 to 27 October, 2021. It should also be noted that formal consent had been taken from each patient before the data collection.

5. Methodology

The study considered the recovery time of a patient from COVID-19 as the survival time. Suppose, T denotes the recovery times (survival times) of the patients suffering from COVID-19. Then the survivorship function, denoted by $S(t)$, is defined as the probability that an individual takes to recover from the COVID-19 more than time t, i.e., suffers up to time t, and is given by,

$$\begin{aligned} S(t) &= P(\text{An individual takes time more than time } t \text{ to recover}) \\ &= P(T > t) \end{aligned} \tag{1}$$

The specific difficulties relating to survival analysis arise largely from the fact that only some individuals have experienced the event and, subsequently, survival times will be unknown for a subset of the study group. For example, collecting information about the recovery times of the COVID patients, some patients may still suffer from the disease. The exact recovery (survival) times of these patients are unknown. As such, these are censored observations or censored times. When the data contains no censored observations, the set of survival times is complete. Here we are using complete data (i.e. no censored observation) with the recovery time of the patients from the COVID-19. In practice, if there are no censored observations, the survivorship function is estimated as the proportion of patients recover after time t is

$$\hat{S}(t) = \frac{\text{Number of patients recovering after time } t}{\text{Total number of patients}} \tag{2}$$

5.1. The association of factors with COVID-19 infection

In the study, we have included three treatment-related factors *viz.* vaccination status (vaccinated or non-vaccinated), place of treatment (Home isolation or COVID care Center/Hospital/Nursing Home), and Pre-disease condition (i.e. whether any illness already exists due to diabetes, blood pressure, etc.), and four demographic factors *viz.* Gender (Male/Female), Age (in years), Location (Rural/Urban), and Food habits (Vegetarian/Non-vegetarian) on the recovery time of the COVID-19 disease. To examine the association of the above-mentioned factors with COVID-19, two non-parametric tests are being used as per their relevance *viz.* Mann–Whitney U Test and Kendall's τ.

5.2. Comparison of recovery times for different risk factors

Suppose $S_1(t)$ and $S_2(t)$ be the recovery functions under two different levels (i.e. symptomatic and asymptomatic) of a factor. To test the null hypothesis that the two levels of the factor have the same recovery function, i.e.,

$$H_0 : S_1(t) = S_2(t)$$

We have used the non-parametric Log-Rank test also called the Mantel-Haenszel test [49]. In this test, the number of recoveries and the

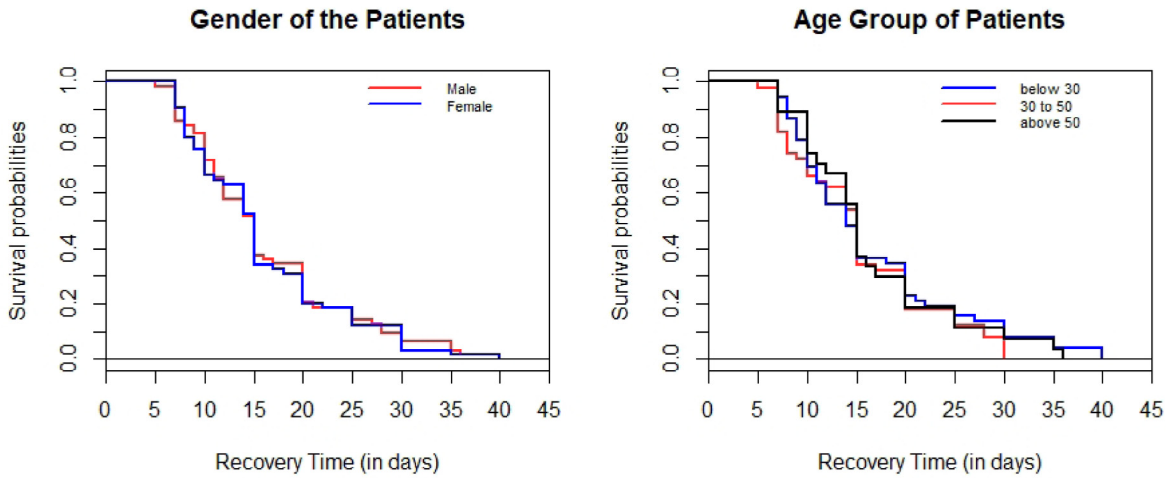

Fig. 1. Recovery curves based on survival functions for risk factors Gender and Age.

Table 1
A (2×2) contingency table for risk levels and recovery status.

	Level 1	Level 2	Total
Recovered	r_{1i}	r_{2i}	$r_{1i} + r_{2i} = r_i$
Not recovered	$n_{1i} - r_{1i}$	$n_{2i} - r_{2i}$	$n_{1i} + n_{2i} - r_{1i} - r_{2i} = n_i - r_i$
	n_{1i}	n_{2i}	n_i

number of patients at risk at each distinct time is observed in a 2×2 contingency table. That is, for the ith recovery time $t_{(i)}$, we may construct a 2×2 contingency table showing the numbers at risk (n_{1i} and n_{2i} for the two levels of the factor) and the number of recoveries (r_{1i} and r_{2i}, respectively). Thus, for the ith recovery time, we have obtained Table 1.

Suppose that the number of recoveries in the two levels of the factor is independent. Then, for fixed values of r_i, n_i, n_{1i} and n_{2i}, the random variable R_i, denoting the entry in the (1, 1) cell of the 2×2 table has hypergeometric distribution with conditional p.m.f.

$$P(R_i = r_{1i}/n_{1i}, n_{2i}, r_i) = \frac{\binom{n_{1i}}{r_{1i}}\binom{n_{2i}}{r_{2i}}}{\binom{n_i}{r_i}}; \quad i = 1, 2, \ldots, k \tag{3}$$

The mean and the variance of the distribution are

$$E_i = E(R_i) = \frac{n_{1i} r_i}{n_i} \quad \text{and} \quad V_i = Var(R_i) = \frac{n_{1i} n_{2i} (r_i - n_i)}{n_i^2 (n_i - 1)}; \quad i = 1, 2, \ldots, k$$

Combining the information from each 2×2 table for each t_i we can define a statistic R as,

$$R = \sum_{i=1}^{k} (R_i - E_i) \tag{4}$$

Which gives an overall measure of the deviation of the observed values of R_i from their expected values. Since $E(R_i) = E_i$, therefore the statistic R has to mean zero, i.e., $E(R) = 0$. Also, since the recoveries are independent of each other, the variance of R is the sum of the variances of R_i, i.e.,

$$Var(R) = \sum_{i=1}^{k} V_i \tag{5}$$

Then, by the large sample property, the variable,

$$U = \frac{R - E(R)}{\sqrt{Var(R)}} \tag{6}$$

is approximately normal, i.e., $U \sim N(0, 1)$. Thus, the test statistic

$$U^2 = \frac{[R - E(R)]^2}{Var(R)} \tag{7}$$

summarize the extent to which the observed recovery times in the two levels of the factor deviate from those expected under the null hypothesis of no differences. Under H_0, the test statistic will have an asymptotic chi-squared distribution with one degree of freedom (Cox, 1972). Thus, to test the hypothesis that the two levels of a factor have the same survivor function, we can carry out a Chi-square test, using the statistic U^2, defined in (4), and this test is known as the *Log-Rank test* or *Mantel-Haenszel test.*

5.3. Impact of risk factors on the recovery time of the COVID-19 patients

The Cox model is a semi-parametric regression model that defines the hazard function $h(t, \boldsymbol{x})$ by a set of p covariates $\boldsymbol{x}^{/} = (x_1, x_2, \ldots, x_p)$ as

$$h(t, \boldsymbol{x}) = h_0(t) e^{\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p} = h_0(t) e^{\boldsymbol{\beta}^{/} \boldsymbol{x}} \tag{8}$$

where t represents the survival time. The term $h_0(t)$ is called the baseline hazard. It corresponds to the value of the hazard if all the x_i are equal to zero, i.e., without any influence of covariates. The coefficients $\boldsymbol{\beta}^{/} = (\beta_1, \beta_2, \ldots, \beta_p)$ measure the impact (i.e., the effect size) of covariates. The logarithm of the hazard ratio in Eq. (8) gives the numerical estimates of the regression coefficients β_i rather than in the shape of $h(t, \boldsymbol{x})$

$$log h(t, \boldsymbol{x}) = log h_0(t) + \boldsymbol{\beta}^{/} \boldsymbol{x}$$

or

$$log \frac{h(t, \boldsymbol{x})}{h_0(t)} = \boldsymbol{\beta}^{/} \boldsymbol{x} = \sum_{i=1}^{p} \beta_i x_i$$

or

$$log(HR_0) = log\left(\frac{\text{Group Hazard}}{\text{Baseline hazard}}\right) = \sum_{i=1}^{p} \beta_i x_i \tag{9}$$

which is a linear function of the covariates $x_1, x_2, \ldots, x_p$.

6. Results and discussion

Based on the methodology discussed above, using Eq. (2) several survivorship functions are depicted under various risk factors *viz*. Gender, Age, Location, Vaccination Status, Pre-disease Condition, Place of Treatments, and Food Habits. A simple comparison of survival functions can be achieved by drawing graphs of the estimated survivorship functions. Figs. 1–4 gives the survival curves for the risk factors included in the study.

After portraying survival functions, the study has tried to identify the factors that are associated with the patients using the Mann–Whitney U test and Kendall's τ. The results of the tests are presented in

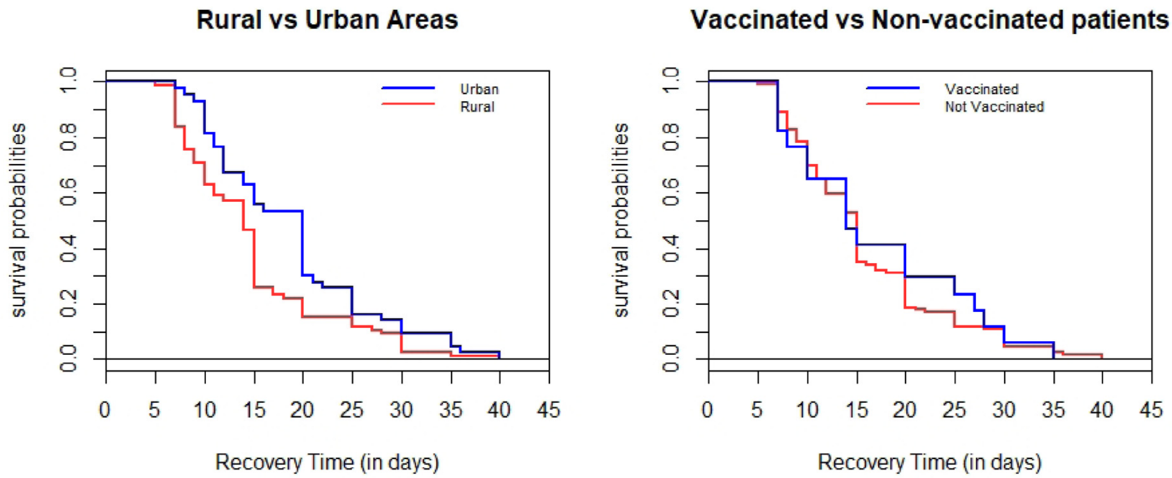

Fig. 2. Recovery curves based on survival functions for risk factors Location and Vaccination Status.

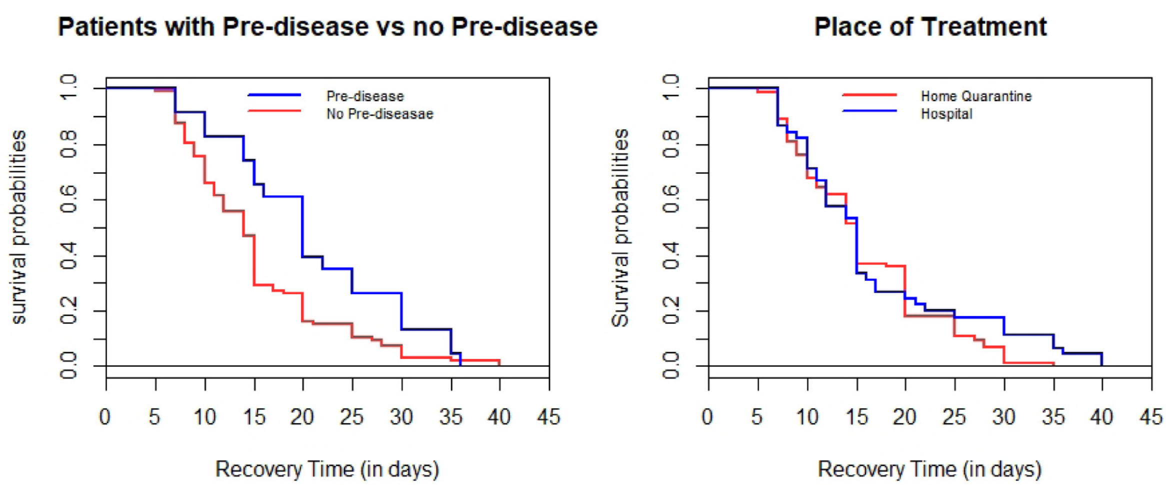

Fig. 3. Recovery curves based on survival functions for risk factors Pre-disease Condition and Place of Treatment.

Table 2. Among all the risk factors considered in the study, only three risk factors are found to be significantly associated with the patients having COVID-19 infections. These three factors are Pre-disease Conditions, Location and Food habit which is also evident from Table 2 as the p-values corresponding to these factors are less than 0.05 (significant at 5%). Moreover, an almost equal proportion of patients of both genders is included in the sample (male: 49.6%, female: 50.4%). It is seen that the mean recovery time for both genders is almost equal, (male: 16.3 days, female: 16.0 days). Regarding age which was not found to be statistically significant, at the time of infection, 38.8% of patients are below age 30 years; 40.3% are of age 30–50 years, and 20.9% of patients are of age 50 years or more.

Interestingly the risk factor vaccination status which is usually considered to be most indispensable to dispose of COVID-19 infection is not statistically significant. Also, at the time of infection, only 13.2% of patients were vaccinated. The median recovery time for the vaccinated and non-vaccinated patients are respectively 14 days and 15 days, which is not statistically significant. This might be because the gap between the day of vaccination and the study period is so close that the effect of vaccination was yet to blossom to its full potential. Again, almost two-thirds of the patients preferred home isolation (Home isolation: 65.1%, COVID care center/hospital: 34.9%). The median recovery time for patients preferring home isolation is 15 days, and that taking treatment in COVID care Center/Hospital/Nursing Home is also 15 days.

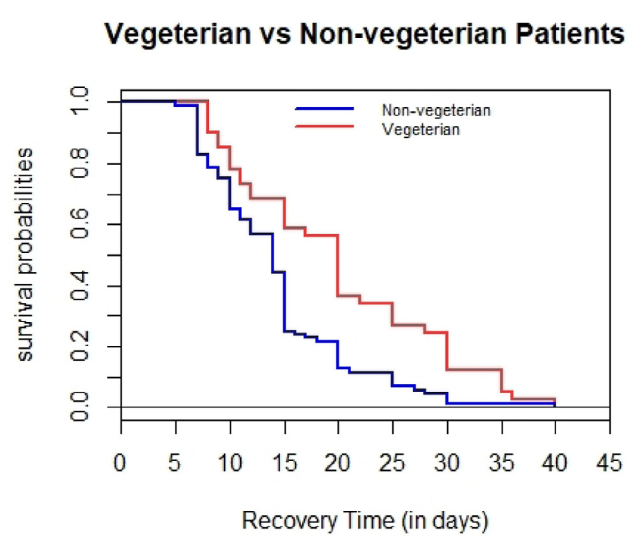

Fig. 4. Recovery curves based on survival function for risk factor Food Habit.

From Fig. 3 of the recovery curves, patients having pre-disease conditions have a longer recovery time than those are not having any

Table 2
Association of Risk factors with COVID-19 Infection.

Risk factors	Status	No. of cases	Percent	Mean Recovery time (MRT)	SD[a] of MRT	Median recovery time	Test statistic	p-value
Vaccination status	Vaccinated	17	13.2%	17.1	9.0	14	Mann–Whitney	0.861
	Not vaccinated	112	86.8%	16.0	7.9	15	$U = 927.0$	
Gender	Male	64	49.6%	16.3	8.2	15	Mann–Whitney	0.874
	Female	65	50.4%	16.0	7.9	15	$U = 2046.0$	
Age	<30	50	38.8%	15.4	7.4	15	Kendall's	0.375
	30–50	52	40.3%	16.6	8.8	14	$\tau = 0.055$	
	≥50	27	20.9%	16.5	7.9	15		
Pre-disease	No	106	82.2%	15.1	7.5	14	Mann–Whitney	0.004
	Yes	23	17.8%	20.7	8.9	20	$U = 748.0$	
Location	Rural	86	66.7%	14.7	7.5	14	Mann–Whitney	0.003
	Urban	43	33.3%	18.8	8.4	20	$U = 1261.5$	
Treatment place	Home	84	65.1%	15.7	7.2	15	Mann–Whitney	0.781
	Hospital	45	34.9%	16.8	9.4	15	$U = 1834$	
Food habit	Veg	41	31.8%	20.0	9.3	20	Mann–Whitney	0.001
	Non-veg	88	68.2%	14.3	6.6	14	$U = 1136.5$	

[a]SD = Standard Deviation.

Table 3
Impact of various factors on the recovery time.

Factors	β	SE[a] of β	e^{β}	CI[a] of e^{β}	z	$P(Z > \|z\|)$
Location	−0.622	0.390	0.537	[0.249, 1.153]	−1.594	0.111
Pre-disease	−1.088	0.474	0.337	[0.133, 0.853]	−2.296	0.022*
Food Habits	0.562	0.257	1.754	[1.059, 2.905]	2.183	0.029*
Location & Pre-disease	1.078	0.548	2.940	[1.005, 8.605]	1.969	0.049*
Location & Food	−0.005	0.443	0.995	[0.418, 2.369]	−0.011	0.991
Pre-disease & Food	0.563	0.524	1.756	[0.628, 4.908]	1.074	0.283
Likelihood ratio test = 26.29 on 6 df, $p < 0.001$						
Wald test = 24.03 on 6 df, $p < 0.001$						
Score (Log-Rank) test = 25.59 on 6 df, $p < 0.001$						

[a]SE = Standard Error; CI = Confidence Interval.

pre-disease condition. The median recovery time for patients with pre-disease is 20 days and no pre-disease is only 14 days. It is because the various pre-disease conditions stance different risks for patients who have suffered COVID-19. The list of such alarming pre-diseases is Diabetes, Cardiovascular disease, Chronic Lung Disease, Hypertension, Obesity, Metabolic disease, etc. The pre-existence of these diseases increases the risk of hospitalization, ICU admission, and even death. Location is another significant risk factor. Patients in urban areas have comparatively longer recovery periods (cf. Fig. 2). One of the key advantages for the patients staying in rural areas is lack of pollution and easier to maintain physical distancing. Usually, urban areas are more contaminated in terms of air and soil compared to rural areas and because of lack of space physical distancing is difficult to maintain. Again, food habit is another significant risk factor, vegetarians have a longer recovery time than non-vegetarians. The median recovery time for vegetarian and non-vegetarian patients is 20 days and 14 days respectively.

From Table 2, the set of covariates that are influencing the recovery time is pre-disease, location, and food habits. Thus, to examine the combined effect of these covariates on the recovery time of the COVID-19 patients, Cox's Proportional Hazard Model [50] is used. The model evaluates simultaneously the effect of several covariates on survival (i.e. recovery time), which can be sophisticatedly adapted in the present context. Table 3 gives the p-values ($Pr(Z > |z|)$) along with the coefficients (β) and its standard errors (SE), exponentiation of coefficients (e^{β}) and its confidence intervals (CI), and critical values (z) of the proportional hazard model.

The p-values for all three tests viz. likelihood ratio test, Wald test, and Score test are significant, indicates that the model is significant, and hence, the omnibus null hypothesis that all of the β's are 0 is rejected. So, we could infer that there is an overall significant relationship between these sets of covariates and recovery time. From the output in the multivariate Cox analysis, the two covariates Pre-disease and Food Habits are significant factors ($p < 0.05$), but the location is found to be an insignificant factor ($p > 0.05$). However, interestingly the interaction between Location and Pre-disease is found to be significant.

The p-value for Pre-disease is 0.022, with a hazard ratio, $HR = e^{\beta} = 0.337$, indicating that the risk of recovery decreases by a factor of 0.337 for the patients having pre-disease conditions if we hold the other covariates constant. That means the recovery time of the patients having pre-disease condition decreases by $(0.337 \times 100 - 100 =)$ 66.3% compared to those without any pre-disease condition. This may be due to the fact that medical practitioners (i.e. doctors, nurses, etc.) are taking special care for those people who are having pre-existing diseases like diabetes, high blood pressure, sugar, etc. and also for these people specific guidelines are there for COVID-19 treatment. However, there is no specific guidelines for the patients those who have no pre-disease condition. Again, the p-value for food habits is 0.029, indicating a significant relationship between food habits and risk of recovery. The vegetarians experience a higher recovery time with an increased hazard by a factor of 1.754 over the non-vegetarian patients. This implies that compared to non-vegetarian patients there is an increase in hazard ratio by 75.4% in the case of vegetarian patients. Moreover, the significant interaction between location and pre-disease condition ($p = 0.049$) signifies that residing in urban areas with having pre-disease conditions is not advantageous with increased hazard by three ($2.94 \sim 3$) times over the patients residing in rural areas with none of the pre-disease conditions. That means urban patients with pre-existing disease need longer recovery time in comparison to the patients residing in rural areas. As air pollution is more in the urban areas compared to rural locations so breathing problems are more acute in urban areas as COVID-19 directly affects human lungs.

7. Conclusion and future scope of the study

The purpose of the current study is to identify the factors that influence the recovery time of the COVID-19 patients with mild symptoms using survival analysis. Here time to recovery of the patient from COVID-19 is the variable of interest. As the recovery time of patients are not normally distributed so non-parametric tests like Mann–Whitney's U test (when only two patient groups are to be compared) and Kendall's τ (for more than two patient groups) were compared. The empirical survival curves of the various patient groups are visualized and compared graphically and also their significant difference were tested using Mantel-Haenszel test. The study identifies that the significant risk factors that prolong the recovery time are Pre-disease, Location, and Food habits. Furthermore, the combined effect of the significant risk factors in the recovery time of the patients is studied using $Cox's$ Regression. The overall findings of the statistical exercise are that the urban patients with pre-disease have the highest recovery time. The vegetarian patients and those living in urban areas also take more time to recover compared to their non-vegetarian and rural counterparts respectively. Other factors or their interaction effects like gender, age and place of treatment are insignificant as per the analysis.

Most studies on the pandemic are based on results from clinical trials. Such studies are basically focused on- to what extent the medical parameters of human body are influenced by COVID-19 infection [51–53]. Some other works focused on the extent in which precautions were followed by people in order to keep them safe from the said infection [54] or to identify the group of people who are more vulnerable to the infection. However, the current paper is focused on the time to recovery of COVID-19 patients and the factors that influence the speed of recovery. The study is based on information collection from the patients. The result of our study could not be compared to other studies because in the existing body of literature, the address of recovery time and its determinants were not addressed. However, some of the findings of our study are in keeping with the popular believes encompassing the COVID-19 like- living in rural area (allows natural physical distancing between people), having comorbidity (increases the severity of the infection), taking protein rich food i.e. non-vegetarian food helps in building resistance against the virus. Here also, we found that such factors affect the rate of recovery from corona infection.

One interesting and unexpected finding of the analysis is that the status of vaccination does not influence the time to recovery. But, at the time of data collection, when the third surge of the Corona virus is on the rampage it is seen that the rate of infection is very rapid but the severity of the virus is minimized to the level of common cold and cough. But the same has not been reflected in the current study. The researchers presume that this might be because the gap between the day of vaccination and the day of data collection in case of most of the subjects are so close that the effect of vaccination was yet to blossom to its full potential. The researchers feel that such studies need to be repeated at a future date to recheck the influence of vaccination.

Declaration of competing 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.

Acknowledgments

The authors would like to thank the Editor-in-Chief, Department Editor and the reviewers for their detailed, careful, and exhaustive comments. These have led to very substantial improvement of the paper.

Informed consent

Informed consent was obtained from each participant.

References

[1] E.E. Werner, R.S. Smith, Overcoming the Odds: High Risk Children from Birth to Adulthood, Cornell University Press, 1992.

[2] N.E. Garmezy, M.E. Rutter, Stress, coping, and development in children, in: Seminar on Stress and Coping in Children, 1979, Ctr for Advanced Study in the Behavioral Sciences, Johns Hopkins University Press, Stanford, CA, US, 1983.

[3] P.J. Haggerty, H.R. Mrazek, Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research, National Academies Press, (US), 1994.

[4] G.K. Schleicher, W. Lowman, G.A. Richards, Case study: a patient with asthma, Covid-19 pneumonia and cytokine release syndrome treated with corticosteroids and tocilizumab, Wits J. Clin. Med. 2 (Si1) (2020) 47–52.

[5] X. Yang, Y. Yu, J. Xu, H. Shu, H. Liu, Y. Wu, L. Zhang, Z. Yu, M. Fang, T. Yu, Y. Wang, 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. 8 (5) (2020) 475–481.

[6] K.S. Yuen, Z.W. Ye, S.Y. Fung, C.P. Chan, D.Y. Jin, SARS-CoV-2 and COVID-19: The most important research questions, Cell Biosci. 10 (1) (2020) 1–5.

[7] A. Hampshire, W. Trender, S.R. Chamberlain, A.E. Jolly, J.E. Grant, F. Patrick, N. Mazibuko, S.C. Williams, J.M. Barnby, P. Hellyer, M.A. Mehta, Cognitive deficits in people who have recovered from COVID-19, EClinicalMedicine 39 (2021) 101044.

[8] O. Ozdemir, Coronavirus disease 2019 (COVID-19): diagnosis and management, Erciyes Med. J. 42 (3) (2020) 242–248.

[9] M. Norouzi, S. Norouzi, A. Ruggiero, M.S. Khan, S. Myers, K. Kavanagh, R. Vemuri, Type-2 diabetes as a risk factor for severe COVID-19 infection, Microorganisms 9 (6) (2021) 1211.

[10] C.V. Ugwueze, B.C. Ezeokpo, B.I. Nnolim, E.A. Agim, N.C. Anikpo, K.E. Onyekachi, COVID-19 and diabetes mellitus: the link and clinical implications, Dubai Diabetes Endocrinol. J. 26 (2) (2020) 69–77.

[11] S.B. Brosnahan, A.H. Jonkman, M.C. Kugler, J.S. Munger, D.A. Kaufman, COVID-19 and respiratory system disorders: current knowledge, future clinical and translational research questions, Arterioscler. Thromb. Vasc. Biol. 40 (11) (2020) 2586–2597.

[12] R.M. Kaplan, A. Milstein, Influence of a COVID-19 vaccine's effectiveness and safety profile on vaccination acceptance, Proc. Natl. Acad. Sci. 118 (10) (2021).

[13] A. Nalbandian, K. Sehgal, A. Gupta, M.V. Madhavan, C. McGroder, J.S. Stevens, J.R. Cook, A.S. Nordvig, D. Shalev, T.S. Sehrawat, N. Ahluwalia, Post-acute COVID-19 syndrome, Nat. Med. 27 (4) (2021) 601–615.

[14] K. Li, G. Chen, H. Hou, Q. Liao, J. Chen, H. Bai, S. Lee, C. Wang, H. Li, L. Cheng, J. Ai, Analysis of sex hormones and menstruation in COVID-19 women of child-bearing age, Reprod. Biomed. Online 42 (1) (2021) 260–267.

[15] A. Duda-Chodak, M. Lukasiewicz, G. Ziec, A. Florkiewicz, A. Filipiak-Florkiewicz, Covid-19 pandemic and food: Present knowledge, risks, consumers fears and safety, Trends Food Sci. Technol. 105 (2020) 145–160.

[16] R. Mahmud, M.M. Rahman, M.A. Rassel, F.B. Monayem, S.J. Sayeed, M.S. Islam, M.M. Islam, Post-COVID-19 syndrome among symptomatic COVID-19 patients: A prospective cohort study in a tertiary care center of Bangladesh, PLoS One 16 (4) (2021) e0249644.

[17] J.R. Kerr, A.L. Freeman, T.M. Marteau, S. van der Linden, Effect of information about COVID-19 vaccine effectiveness and side effects on behavioural intentions: two online experiments, Vaccines 9 (4) (2021) 379.

[18] M. Augustin, P. Schommers, M. Stecher, F. Dewald, L. Gieselmann, H. Gruell, C. Horn, K. Vanshylla, V. Di Cristanziano, L. Osebold, M. Roventa, Post-COVID syndrome in non-hospitalised patients with COVID-19: a longitudinal prospective cohort study, Lancet Reg. Health-Europe 6 (2021) 100122.

[19] M.P. Sormani, N. De Rossi, I. Schiavetti, L. Carmisciano, C. Cordioli, L. Moiola, M. Radaelli, P. Immovilli, M. Capobianco, M. Trojano, P. Zaratin, Disease-modifying therapies and coronavirus disease 2019 severity in multiple sclerosis, Ann. Neurol. 89 (4) (2021) 780–789.

[20] S. Simpson-Yap, E. De Brouwer, T. Kalincik, N. Rijke, J.A. Hillert, C. Walton, G. Edan, Y. Moreau, T. Spelman, L. Geys, T. Parciak, Associations of disease-modifying therapies with COVID-19 severity in multiple sclerosis, Neurology 97 (19) (2021) e1870–85.

[21] F. Safavi, B. Nourbakhsh, A.R. Azimi, B-cell depleting therapies may affect susceptibility to acute respiratory illness among patients with multiple sclerosis during the early COVID-19 epidemic in Iran, Mult. Scler. Relat. Disord. 43 (2020) 102195.

[22] A. Czarnowska, W. Brola, O. Zajkowska, S. Rusek, M. Adamczyk-Sowa, K. Kubicka-Bączyk, A. Kalinowska-Łyszczarz, K. Kania, A. Słowik, M. Wnuk, M. Marona, Clinical course and outcome of SARS-CoV-2 infection in multiple sclerosis patients treated with disease-modifying therapies—the Polish experience, Neurol. Neurochir. Polska 55 (2) (2021) 212–222.

[23] M. Goumenou, D. Sarigiannis, A. Tsatsakis, O. Anesti, A.O. Docea, D. Petrakis, D. Tsoukalas, R. Kostoff, V. Rakitskii, D.A. Spandidos, M. Aschner, COVID-19 in Northern Italy: An integrative overview of factors possibly influencing the sharp increase of the outbreak, Mol. Med. Rep. 22 (1) (2020) 20–32.

[24] A.D. Smethers, B.J. Rolls, Dietary management of obesity: cornerstones of healthy eating patterns, Med. Clin. North Am. 102 (1) (2018) 107–124.

[25] C. Wu, X. Chen, Y. Cai, X. Zhou, S. Xu, H. Huang, L. Zhang, X. Zhou, C. Du, Y. Zhang, J. Song, Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China, JAMA Internal Med. 180 (7) (2020) 934–943.

[26] S.R. Bornstein, R. Dalan, D. Hopkins, G. Mingrone, B.O. Boehm, Endocrine and metabolic link to coronavirus infection, Nat. Rev. Endocrinol. 16 (6) (2020) 297–298.

[27] J.M. O'Brien Jr., G.S. Phillips, N.A. Ali, M. Lucarelli, C.B. Marsh, S. Lemeshow, Body mass index is independently associated with hospital mortality in mechanically ventilated adults with acute lung injury, Crit. Care Med. 34 (3) (2006) 738.

[28] D. Petrakis, D. Margina, K. Tsarouhas, F. Tekos, M. Stan, D. Nikitovic, D. Kouretas, D.A. Spandidos, A. Tsatsakis, Obesity-a risk factor for increased COVID-19 prevalence, severity and lethality, Mol. Med. Rep. 22 (1) (2020) 9–19.

[29] A. Sanyaolu, C. Okorie, A. Marinkovic, R. Patidar, K. Younis, P. Desai, Z. Hosein, I. Padda, J. Mangat, M. Altaf, A, SN Compr. Clin. Med. 2 (2020) 1069–1076.

[30] H.J. Han, C. Nwagwu, O. Anyim, C. Ekweremadu, S. Kim, COVID-19 and cancer: From basic mechanisms to vaccine development using nanotechnology, Int. Immunopharmacol. 90 (2021) 107247.

[31] L. Derosa, C. Melenotte, F. Griscelli, B. Gachot, A. Marabelle, G. Kroemer, L. Zitvogel, The immuno-oncological challenge of COVID-19, Nature Cancer 1 (10) (2020) 946–964.

[32] B. Li, J. Yang, F. Zhao, L. Zhi, X. Wang, L. Liu, Z. Bi, Y. Zhao, Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China, Clin. Res. Cardiol. 109 (5) (2020) 531–538.

[33] S. Shi, M. Qin, B. Shen, Y. Cai, T. Liu, F. Yang, W. Gong, X. Liu, J. Liang, Q. Zhao, H. Huang, Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China, JAMA Cardiol. 5 (7) (2020) 802–810.

[34] F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, L. Guan, Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study, Lancet 395 (10229) (2020) 1054–1062.

[35] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet 395 (10223) (2020) 497–506.

[36] T. Chen, D.I. Wu, H. Chen, W. Yan, D. Yang, G. Chen, K. Ma, D. Xu, H. Yu, H. Wang, T. Wang, Clinical characteristics of 113 deceased patients with coronavirus disease 2019: retrospective study, Bmj 368 (2020).

[37] J. Xiong, O. Lipsitz, F. Nasri, L.M. Lui, H. Gill, L. Phan, D. Chen-Li, M. Iacobucci, R. Ho, A. Majeed, R.S. McIntyre, Impact of COVID-19 pandemic on mental health in the general population: A systematic review, J. Affect. Disord. 277 (2020) 55–64.

[38] R.S. Soumya, T.G. Unni, K.G. Raghu, Impact of COVID-19 on the cardiovascular system: a review of available reports, Cardiovasc. Drugs Ther. 35 (3) (2021) 411–425.

[39] G. Lippi, J. Wong, B.M. Henry, Hypertension in patients with coronavirus disease 2019 (COVID-19): a pooled analysis, Pol. Arch. Intern. Med. 130 (2020) 304–309.

[40] P. Zhang, L. Zhu, J. Cai, F. Lei, J.J. Qin, J. Xie, Y.M. Liu, Y.C. Zhao, X. Huang, L. Lin, M. Xia, Association of inpatient use of angiotensin-converting enzyme inhibitors and angiotensin II receptor blockers with mortality among patients with hypertension hospitalized with COVID-19, Circ. Res. 126 (12) (2020) 1671–1681.

[41] S. Spudich, A. Nath, Nervous system consequences of COVID-19, Science 375 (6578) (2022) 267–269.

[42] S. Nannoni, R. de Groot, S. Bell, H.S. Markus, Stroke in COVID-19: a systematic review and meta-analysis, Int. J. Stroke 16 (2) (2021) 137–149.

[43] Q. Ruan, K. Yang, W. Wang, L. Jiang, J. Song, Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China, Intensive Care Med. 46 (5) (2020) 846–848.

[44] O. D'Ecclesiis, C. Gavioli, C. Martinoli, S. Raimondi, S. Chiocca, C. Miccolo, P. Bossi, D. Cortinovis, F. Chiaradonna, R. Palorini, F. Faciotti, Vitamin D and SARS-CoV2 infection, severity and mortality: A systematic review and meta-analysis, PLoS One 17 (7) (2022) e0268396.

[45] M. Peng, J. He, Y. Xue, X. Yang, S. Liu, Z. Gong, Role of hypertension on the severity of COVID-19: A review, J. Cardiovasc. Pharmacol. 78 (5) (2021) e648.

[46] D. Liu, X. Zeng, Z. Ding, F. Lv, J.L. Mehta, X. Wang, Adverse cardiovascular effects of anti-COVID-19 drugs, Front. Pharmacol. 2250 (2021).

[47] M. Gamerra, E.D. Corso, E. Cantone, COVID-19: the crucial role of the nose, Braz. J. Otorhinolaryngol. 87 (2021) 118–119.

[48] J. Matias, G. Campanari, G. Souza, V. Lima, R. Tofano, C. Detregiachi, S. Barbalho, Metabolic syndrome and COVID-19, AIMS Bioeng. 7 (4) (2020) 242–253, http://dx.doi.org/10.3934/bioeng.2020021.

[49] N. Mantel, W. Haenszel, Statistical aspects of the analysis of data from retrospective studies of disease, J. Natl. Cancer Inst. 22 (4) (1959) 719–748.

[50] D.R. Cox, Regression models and life tables (with discussion), J. R. Stat. Soc. B34 (1972) 187–220.

[51] J.A. Malik, S. Ahmed, M. Shinde, M.H. Al-Marmash, S. Alghamdi, A. Hussain, S. Anwar, The impact of COVID-19 on the comorbidities: a review of recent updates for combating it, Saudi J. Biol. Sci. 29 (2022) 3586–3599.

[52] M. Hada, A.D. Mosholder, K. Leishear, S. Perez-Vilar, Systematic review of risk of SARS-CoV-2 infection and severity of COVID-19 with therapies approved to treat multiple sclerosis, Neurol. Sci. (2022) 1.

[53] P. Brodin, Immune determinants of COVID-19 disease presentation and severity, Nat. Med. 27 (1) (2021) 28–33.

[54] P.G. Velden, M. Marchand, B. Cuelenaere, M. Das, Pre-outbreak determinants of perceived risks of corona infection and preventive measures taken. A prospective population-based study, Plos One 15 (7) (2020) e0234600.