

## Covid-19 Outcome Prediction Model by Using Radio-Diagnostic Methods

Dragana Mogilevska-Gruevska  

Radiologist, PHI - General Hospital – Kochani, North Macedonia

Ilija Gruevski 

Associate Professor, Faculty of economics, Goce Delcev University, Stip, North Macedonia

Golubinka Boshevska 

Professor of Microbiology, Faculty of Medical Sciences, Goce Delcev University, Stip, North Macedonia

Klime Gjoreski

Professor of Radiology, Faculty of Medicine St. Cyril end Methodius, Skopje, North Macedonia

### Abstract

**Research goal:** The goal of this research is to create a prediction model for a possible outcome (death or survival) of COVID-19, which model incorporates the easily available radio diagnostic methods such as classical radiology and the disease scoring system. Part of the goal of this study is to define the chances and probabilities of occurrence of death as a result of the primary disease and to identify the risk factors that have the highest influence on the final outcome of COVID-19. **Methodology:** The methodological approach used in this study is the *binary logistic regression* which is part of the group of generalized linear statistical methods. **Results:** Results show that patients with complications and comorbidities have the highest chances of death from COVID-19 (OR 16,53 with CI 8,21 - 33,25 and 4,08 with CI 1,34 - 12,38). Men are also exposed to higher but insignificant mortality risk with OR 1,55 with CI 0,86 - 2,80. Every additional year of age increases the mortality risks by 1,06 times (CI 1,03 - 1,09), while every additional score of the primary disease leads to increased chances of unwanted outcome by 1,24 (CI 1,04 - 1,47). **Conclusion:** The mortality outcome of COVID-19 is not an exclusive consequence of the primary disease but it is usually determined in correlation with different comorbidities and existing complications as well as other standard influencing factors such as age and gender. **Contribution and significance of the research:** The primary importance of this research is the fact that it allows for an improved precision and upgrade to the basic model of standard factors by using new predictors, specifically secondary complications from the radio-graphic picture and scoring of the primary disease, which leads to higher utilization of cheap and easily available radio-diagnostic methods.

### Introduction

#### Background

The colossal dimensions of the pandemics caused by the virus SARS-CoV-2 imposed the need to study the associated risks of the disease that it causes. Namely, the price paid was measured in human lives, and until March 12th 2024 COVID-19, caused by the corona virus, has claimed 7.004 732 human lives. In Macedonia, the corresponding number is 9981 deaths [1]. A few

fundamental factors that influence the final outcome of COVID-19 can be singled out from the wider literature and those are generally related to the differences in *gender* and *age* as well as the presence or absence of *comorbidities*. So, for example, the evidence related to the *gender* as a determining factor is indisputable. Numerous studies confirm the findings that the male infected population is actually more susceptible to death in comparison to the female one [2]. For

### More Information

#### How to cite this article:

Mogilevska-Gruevska D, Gruevski I, Boshevska G, Gjoreski K. Covid-19 Outcome Prediction Model by Using Radio-Diagnostic Methods. Eur J Med Health Res, 2024;2(3):199-211.

DOI: 10.59324/ejmhr.2024.2(3).24

#### Keywords:

COVID-19,  
logistic regression,  
radio-diagnostic methods,  
scoring system,  
North Macedonia.



This work is licensed under a Creative Commons Attribution 4.0 International License. The license permits unrestricted use, distribution, and reproduction in any medium, on the condition that users give exact credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if they made any changes.

example, according to a meta-analysis of 3.111.714 globally recorded cases of COVID-19, mortality rates were higher in the male population by 1,39 times compared to the mortality rates among the female population [3]. Medical studies point out few possible reasons for the difference in gender related mortality rates, some of which are: the emphasized genetic expression of ACE2 receptor in men [4], women hormonal status [5], characteristics of immunity system of men [6], higher levels of testosterone [7], higher predisposition of chronic diseases and hazardous behavior of men [8], etc. Or course, literature is abundant with more reserved, inconclusive studies related to the previous hypothesis, i.e. gender difference variations when it comes to mortality is found solely within the different age groups, but not in the whole sample [9]. Age is considered as another basic criteria defining the outcome of COVID-19. In according to this, great number of studies explore the correlation between the age groups and the disease in the context of mutual multifactorial influence with the other determining factors which are tightly related to the aging population such as chronic diseases and comorbidities. After eliminating the co-founding effects of the other variables, great number of these studies conclude that age as an independent factor is closely related to the mortality rates of COVID-19. For example, the results of a meta-study held on 470.034 participants show that those above 75 years of age had 13 times higher risk of death compared to those under 65 years of age [10]. Within the same study, the higher death rate was partially explained in relation to other risk factors (low forced expiratory volume, high systolic blood pressure, weak hand grip, multiple chronic conditions, etc.), while the "healthy" elderly people had significantly lower risk of death when they were infected with the corona virus. Still, researchers concluded that the age as an independent risk factor is positively related to higher risks of unwanted outcomes. *Comorbidities* are the third and probably the most researched factor of influence related to death outcomes of COVID-19. Comorbidity is the simultaneous presence of one or more diseases or medical conditions, usually long lasting or chronic in a patient along with the primary disease (in this case COVID-19). A USA study of 600.241 death cases related to COVID-19 shows highest prevalence to hypertension (19,6 %), diabetes mellitus (15,9 %), ischemic heart disease (10,9%), heart weakness i.e. cardiomyopathy (7,7%), heart arrhythmia (7,5%), other circulatory system diseases (6,6%), cerebrovascular diseases (5%) and obesity (4,1%) [11]. From 14 studies reviewed under a meta-analysis, it was concluded that the incidence of disseminated intravascular coagulation - DIC was 3% among the COVID-19 affected patients. In addition to this, death cases caused by COVID-19 were

positively associated to the presence of DIC with 2,46 higher chances of death in these patients compared to the patients without coagulopathy [12]. Kidney diseases, especially chronic kidney disease or CKD are also related to the risk of death in case of COVID-19 infection. The prevalence of death was 1,42 times higher in patients with chronic kidney disease compared to patients without it, while when patients had comorbidities in combination with CKD, the highest mortality rate was manifested in those with diabetes mellitus and hypertension [13]. It is worth mentioning the pulmonary diseases such as chronic obstructive pulmonary disease - COPD and asthma (pulmonary pneumonias in this study are classified as complications). So, patients with COPD had higher rates of hospitalization compared to patients without this disease (62% compared to 28%), while the chances of death outcome by COVID-19 were 2,1 times higher in the former than in the latter [14]. When it comes to asthma, the conclusion of its influence on the mortality rate of patients with COVID-19 is inconclusive with many studies that fail to determine a correlation and less studies that actually confirm a relation to the higher risk of death. The studies, so far, do not classify asthma in the group of top 10 comorbidities associated with unwanted outcome or fatality from COVID-19 [15]. When it comes to malignant diseases as comorbidity, one big study of 1.050.045 patients (out of which 27.760 had any type of cancer) reveals that the mortality rate in these patients was 17,58 % compared to 11% in those patients without malignant diseases. After an adjustment was made in logistic regression, the COVID-19 patients with any type of cancer had 1,21 times higher chances of death in hospital conditions caused by any reason, in comparison to those patients without cancer, while the chances of developing acute respiratory failure were 1,14 times higher. Novelty in this research is the addition of new independent determinants to the basic model which is composed of fundamental factors in order to see if they can also be used as adequate predictors which can improve the precision in foreseeing the associated risks with COVID-19. The idea is to use radio-diagnostic methods, that is, the instruments of radiology as a cheap, quick and easily available diagnostic approach. In this context, the new determinants that we propose refer to the complications and scoring system. The first determinant which is marked as *complications*, actually refers to the pulmonary complications or to be more precise - bronchopneumonia, pneumonia and pleural effusion. These conditions appear as secondary as a result of the commonly developed bacterial superinfection of the organism due to weakening of the immune system by the primary interstitial pulmonary pneumonia which is caused by the SARS-CoV-2 virus itself. Pleural effusion as a type of pulmonary



complication is actually accumulation of fluids between the parietal and visceral pleura (or pleural cavity). This could be a result of a disease of the surrounding parenchyma such as infection, heart weakness, pulmonary embolism, malignancy or inflammation, and in rare cases it can occur spontaneously by itself without any manifestation of a pre-existing disease. The **scoring system** as a separate diagnostic method, as it was already explained, is used to rank the weight and intensity of the clinical i.e. radiological manifestation of the atypical and characteristic pulmonary pneumonia which is a direct cause of the corona virus itself. This method contains 9 scores or stages which are aligned by the intensity and the degree of lung tissue involvement starting with the radiological image on the lowest score 0 (negative radiological image) up to the highest score 8 (almost total involvement of both lungs). According to the methodology of “*Radiographic assessment of edema score*”, depending on the involvement of the lung parenchyma, the grading is done by using a scale 0-4 for the right lung and 0-4 for the left lung in the following way:

**0** - without involvement of the lung parenchyma

**1** - involvement of the lung parenchyma  $\leq 25\%$

**2** - involvement of the lung parenchyma from 25% to 50%

**3** - involvement of the lung parenchyma from 50% to 75%

**4** - involvement of the lung parenchyma  $\geq 75\%$

Total score is achieved by summing up the values of both lungs with the possibility to grade from 0-8. We should mention that if a correlation is detected between these two determinants and the mortality rate of COVID-19, it does not mean that there is a causal relation between them automatically. Since the transmission mechanisms may be from another origin, they cannot be considered as a direct cause for such condition.

#### Research aim

The aim of this research is to create a model of possible outcome (death or survival) of COVID-19, which incorporates the easily available diagnostic methods such as classical radiology and scoring system of grading the disease. Part of the aim of this study is to define the chances and probabilities of mortality by the primary disease and so to identify the risk factors which have the highest influence to the outcome of COVID-19.

#### Contribution and significance of the research

The primary importance of this research is that it allows for an improved precision and upgrade to the basic model based on standard influencing factors with new non- standard predictors of the disease, specifically secondary complications from the radio-graphic picture and scoring of the primary disease, which leads to higher utilization of cheap and easily available radio-diagnostic methods. This model generates the chances

and probabilities of death outcome of COVID-19 in any individual patient, and this in turn allows for individual risk assessment of death by the disease which will encourage activation of special protocol for treatment. Apart from getting the coefficients of determination of the individual factors, this model also portrays the co-integrated effects of two or more factors towards the final outcome of the disease (for example gender and age or gender, score and comorbidity, etc.). It is worth mentioning the adaptability of the model to a wider use in risk assessments from other diseases and also its use in other similar studies in the field of clinical medicine and pharmacology. However, the focus of this paper is to explain its potential application in the above mentioned cases related to the COVID-19 disease, which in a way was a medical “novel” in the field of epidemiology.

#### Reasons for choosing the subject of research

The reasons for choosing to study the COVID-19 are more of a practical nature. First, it is the availability of data of infected patients who came for an X-ray in the PHI - General Hospital - Kochani, which formed a sample in total of 1013 patients. We have to mention that all ethical considerations were met during the sampling process, and approval was asked and received from the patients themselves or their families in order to be able to use the data. The second reason is the availability and the experience of applying diagnostic methods radiography and scoring system to grade the scale of the disease.

#### Used Methodology

The methodological approach used in this research is the *binary logistic regression* which is part of the group of generalized linear statistical models. Its use in medical research is exceptionally large, therefore it is being used extensively in clinical research, in the field of general and family medicine, pharmacology and other fields, i.e. everywhere there is a need for classification, explanation (finding a correlation) or predicting values by some characteristics, conditions or outcomes (for example death or survival, etc.). The end result of the model are the so called *odd ratios* which point to an increased, lowered or steady chances of categorization of a certain observation (for example, a patient with COVID-19) into one of the two outcome categories (death or survival). Apart from the chances of a certain outcome, the model precisely defines the probability of the specific outcome which can be from 0 (definite non-occurrence of the observed event) and 1 (definite occurrence of the observed event). The model contains two or more variables out of which only one is dependent variable, while one or more are independent variables. The dependent variable contains only binary or dichotomous values for example: 1 (death, getting a cancer, getting a heart attack, etc.) and 0 (survival, no cancer, no heart attack,



etc.). The independent variables (predictors, determinants or influencing factors) in the model could be one or more variables with dichotomous (binary), ordinal (categorical) and continuous values. For example, if the aim is to define the predisposition of male/female population towards heart attack then binary values could be used (1 for male and 0 for female). While if it is important to see how age, glycemc index value or the blood pressure influences the heart attack, then continuous variables will be used.

**Defining the model of logistic regression**

Considering the above mentioned elements and characteristics of logistic regression, we can say that it has the following expression (1):

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (1)$$

The figures on the left side  $\ln(p/(1-p))$  is actually logarithmic value of the chances of a specific event or *log (odds) ratio* and the same serves as a dependent variable,  $b_0$  is the intercept,  $b_1, b_2$  and  $b_n$  are the regression coefficients of the independent variables and  $x_1, x_2$  and  $x_n$  are the raw values of the independent variables. From this we conclude that the results of the regression are actually the regression coefficients  $b_1 \dots b_n$ , which express the percentage change of the logarithmic value of odds, caused by a unit change of the independent variables  $x_1 \dots x_n$ . However, the above coefficients do not express the *percentage change of odds* but they express the *percentage change of the logarithmic value of odds* of a certain event, and those are not identical. Fortunately, there is an inherent relation between them which comes from the relation between logarithmic and exponential function. Namely, the logarithmic value of any number X is equal to the basis of the natural logarithm with exponent from the same number X i.e.  $\ln(X) = \exp(X)$ . Thus, if we want to get to the odds of a possible outcome (or odds ratio) as a result of the impact of separate factors  $x_1 \dots x_n$ , we can do it simply by transforming the obtained coefficients  $b_1 \dots b_n$ , through using the following template (2):

$$OR(b_i) = \exp(b_i) \quad (2)$$

If there is a need to estimate the of outcome odds from the co-integrated (joint) influence of more factors at the same time (for example factors  $x_1$  and  $x_2$ ), the template would look like this (3):

$$OR(b_1, b_2) = \exp(b_1 + b_2) \quad (3)$$

If the aim is to estimate the outcome odds of *the control or referent group*, then the same template can be used to transform the intercept  $b_0$  (4):

$$OR(b_0) = \exp(b_0) \quad (4)$$

Finally, if we want to express the odds of manifesting a certain event (odds event) which is influenced by certain factors  $x_1 \dots x_n$ , and whose original coefficients are  $b_1 \dots b_n$ , then we need to use the following template (5):

$$Odds(e) = \exp(b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n) \quad (5)$$

Recently, we demonstrated the capabilities of logistic regression from the aspect of determination of the chances of a certain outcome or event i.e. ORs, but from this statistical model we can also derive the probabilities of a certain event or, in other words, we can predict or forecast a certain medical condition and also measure the risk associated with it. The template which can be used to estimate the probabilities of an observed event (for example: death by COVID-19) is shown with the next formula (6):

$$p(e) = \frac{odds(e)}{1+odds(e)} \quad (6)$$

The relationship between the probability and the odds concepts is more than obvious, as it can be seen in the above presentation.

**Sample Characteristics (Descriptive Statistics)**

The sample in our research is comprised of 1013 COVID-19 positive patients, confirmed by Polymerase Chain Reaction (PCR) and a positive clinical picture. As part of the whole population, the sample covers the north-eastern region of the Republic of North Macedonia, taken into consideration that these patients were sent to the radio-diagnostic department of the PHI - General Hospital - Kochani in the period between September 2020 and September 2022. Analogous to the time frame, we can conclude that Wuhan, British and Delta variants of the virus SARS-CoV-2 were circulating during the observed period. In continuation we present the characteristic of the sample in short, which have been processed and presented in detail in the following table No.1.



**Table 1: Sample characteristics (descriptive statistics) (N=1.013 patients)**

Characteristics	Category	Died (%)	Survived (%)	N=1013; (%)
<b>COVID-19 outcome</b>		77 (7,60%)	936 (92,40%)	1013 (100%)
<b>Gender</b>	Male	51 (10,20%)	500 (89,8%)	551 (100%)
	Female	26 (5,96%)	436 (94,04)	462 (100%)
<b>Age (in years)</b>	Average	71,6	55,7	56,9
	Median	72	58	59
	IQR	(65-78)	(45-67)	(46-68)
	Minimal	47	16	16
	Maximal	88	91	91
	Interval	41	75	75
<b>Age Groups</b>	0-20 years	0 (0%)	9 (100%)	9 (100%)
	21-30 years	0 (0%)	42 (100%)	42 (100%)
	31-40 years	0 (0%)	122 (100%)	122 (100%)
	41-50 years	1 (0,62%)	159 (99,38%)	160 (100%)
	51-60 years	7 (3,52%)	192 (96,48%)	199 (100%)
	61-70 years	24 (8,30%)	265 (91,70%)	289 (100%)
	71-80 years	30 (20,83%)	114 (79,17%)	144 (100%)
	81-90 years	15 (32,61%)	31 (67,39%)	46 (100%)
	91-100 years	0 (0%)	2 ( 100%)	2 (100%)
	<b>Complications</b>	With	66 (31,43%)	144 (68,57%)
Without		11 (1,37%)	792 (98,63%)	803 (100%)
<b>Score</b>	Median	4	2	2
	IQR	(2-5)	(1-3)	(1-3)
	Minimal	1	0	0
	Maximal	8	8	8
	Interval	7	8	8
<b>Intensity (grade)</b>	Score 0	0 (0%)	185 (100%)	185 (100%)
	Score 1	8 (3,32%)	233 (96,68%)	241 (100%)
	Score 2	15 (6,28%)	224 (93,72%)	239 (100%)
	Score 3	10 (7,63%)	121 (92,37%)	131 (100%)
	Score 4	12 (12,37%)	85 (87,63%)	97 (100%)
	Score 5	14 (20,59%)	54 (79,11%)	68 (100%)
	Score 6	13 (38,24%)	21 (61,76%)	34 (100%)
	Score 7	4 (26,67%)	11 (73,33%)	15 (100%)
	Score 8	1 (33,34%)	2 (66,66%)	3 (100%)
<b>Comorbidities</b>	With	73 (12,37%)	517 (87,63%)	590 (100%)
	Without	4 (0,95%)	419 (99,05%)	423 (100%)

Source: Author’s calculations

The table from above shows that the total (overall) rate of mortality of 7,60% is unequally dispersed among both genders, with higher prevalence of mortality by COVID-19 in the male population with 51 death cases, while there are 26 death cases in the female population. This means that the mortality rate as a ratio

between the number of deceased and the total number of infected male patients was 10,20 %, while the same was 5,96% in the female population. If we analyze the age characteristics from the mortality aspect, we can conclude that the average age of deceased patients was 71.6 years (median 72 years) with an interquartile rank



from 65 to 78 years. The minimum age of the youngest deceased patient was 47 while the maximum age of the oldest deceased patient was 88 with an interval of 41 years. The distribution of deceased people according to age groups has an inclination tendency starting with the age group 41-50 years old with its peak at the age group 71-80 years old. In terms of mortality rate, the most vulnerable group was the one from 81 - 90 years old with a death rate of up to 32,61%. The numbers of deaths with complications, which were registered in 66 deceased people, are also striking. Similar conclusion can be drawn for the comorbidities, which were present in 73 deceased patients. Score 6 stands out as the most critical score with 38,24% mortality rate, as well as score 8 with 33,34% rate. The presence of deaths in the lower ranks refers to the conclusion that the death rate of COVID-19 is not an exclusive consequence of the primary disease but is more often a synergistic result of the influence of the other factors, such as secondary disease or comorbidity.

**Main model of the Research**

This part contains the main model of the research which was developed according to the methodology of logistic regression explained above. Specifically, the presented model has characteristics of the binary logistic regression, considering the binary nature of the dependent variable. According to this, if we start from the general form of the model presented in formula (1), we will get the following form (7):

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1Sex + b_2Age + b_3Complications + b_4Score + b_5Comorbidity \tag{7}$$

Here,  $\ln(p/(1-p))$  is the logarithmic value of chances of death from COVID-19, *sex*, *age*, *complications*, *score* and *comorbidity* are the independent determinants (or factors of influence),  $b_1$ ,  $b_2$ ,  $b_3$ ,  $b_4$ ,  $b_5$  are the corresponding coefficients while  $b_0$  is the corresponding coefficient of the reference group or *the intercept*. The obtained coefficients express the percentage change of the logarithmic value of the death odds by COVID-19 or log (odds) ratio as a result of a unit change of the independent variables. According to this, the model contains 4 binary variables out of which one is dependent and it is the *death by COVID-19* while the other three are independent: *gender*, *complications* and *comorbidity*. They contain *dichotomous values* in their data series, so the event of death, male gender, complications and comorbidity in the registered patient is marked by 1, while in the opposite case it is used 0. The model incorporates two independent variables out of which one is *continuous* while the other one is *categorical*. The variable with continuous data is the *age* expressed in years, which values span from 16 in the case of the youngest patient up to 91 in the oldest

patient. When it comes to the *score* variable, the values range from 0 for the lowest intensity of the disease up to 8 for the highest intensity of the disease. Actually, the score variable has features of a categorical variable given that its purpose is to categorize or grade the intensity of the disease. If we take this into consideration, we can conclude that the *control* or *reference group* is made up of *female patients, with age 0 and score 0, without any comorbidity and at the same time do not present any complications on the lung x-ray.*

**Results**

The calculations related to the results of the logistic regression are done with the help of sophisticated extension of the *Excel* software package, named *Xrealstats* and it delivers the results based on the raw data using the so called *Newton method* of calculations with iterations. Another advantage of the additional software is that it directly derives the odds of a certain event or *the odds ratio* which are the main focus of the medical research. In addition, this version has the possibility to derive calculations for predicting the *probability of a certain event* (in this case death by COVID-19). The software package *Xrealstats* can be freely downloaded from the website <https://real-statistics.com/free-download/real-statistics-resource-pack> and upgraded to any basic version of *Excel*. The obtained results are presented in picture No.1 which is taken from the *Excel* program.

**Chances of death (odds ratio)**

Coefficients of the logarithmic value of chances of death or *log (odds) ratios* are presented in the upper left corner, as well as in the lower table in the right side of the image and they are marked by *coeff b*. Coefficients of the chances of death or *odds ratios* are marked by *exp(b)* in the same table, and they are derived from the previous coefficients according to the expression (2) and (4). According to this, the odds of death by COVID-19 are 1,55 times higher in male population compared to female population with confidence interval from 0,86 to 2,80 (CI from 0,86 to 2,80), if we assume that the other factors remain fixed or unchanged. In terms of the other determinants, we can say that COVID death odds are 16,53 times higher in patients with complications compared to those without complications (CI from 8,21 to 33,25), while patients with comorbidity had 4,08 times higher death odds compared to those without comorbidity (CI from 1,34 to 12,38). Furthermore, for every additional year of age, the death odds increases 1,06 times (CI from 1,03 to 1,09) while with every additional rank or score of the disease the chances of death from COVID-19 also increases 1,24 times (CI from 1,04 to 1,47). For example, if the patient is 50 years old, his/her death odds are 18,42 times higher compared to patients with 0 years (control group) or  $(1,06)^{50}$ . If the patient has score 5, then his/her death odds are 2,93 times higher



compared to a patient from the control group or (1,24)<sup>5</sup>. As we can see, the increase of death odds in continuous variables is expressed multiplicatively [17]. When it comes to the death odds ratio in patients from the control group, they are determined by the value of the intercept  $exp(b_0)$  and according to our model it is 6.33E-05, or 0.0000633. Inversely, it can be interpreted that the chances of survival of the control group patients are 99.99367% (1-0.0000633). From the values of the CI and the p-value, we can conclude that the coefficients of all determinants are statistically significant ( $p < 0.05$ ) except the gender, whose CI contains the value 1 and accordingly, in this case there is not enough evidence to reject the null hypothesis.

**Cointegration of factors**

The model allows for calculation of the odds from the integrated impact of two or more factors simultaneously (*cointegration of factors*). For example,

if the patient is male with comorbidity and complications, the estimated death odds would be 103,86 times higher compared to those from the control group (1,55\*4,04\*16,53) or alternatively  $exp(0,44+2,80+1,40)$ . If we, again, assume that the patient is male with comorbidity and complications, he is 50 years old with score 5, then the death odds would be 5.641,85 times higher compared to the control group (1,55\*4,08\*16,53\* 18,42\*2,93). In this case, the chances of death are measured 0,357 (0,0000633\* 5.641,85), which means there are 35,7% chances that the patient with these characteristics would die from the primary disease. Analogous and inversely, this coefficient can be interpreted as following: the survival chances of male patients who are 50 years old, with comorbidity, have complications and score 5 on their X-ray picture are only 64.3% (1-0,357).

Coeff	LL0	-272,415	Covariance Matrix						Converge	
	LL1	-160,24	1,156305	-0,04535	-0,0122	-0,07224	-0,01699	-0,15407	-5,4E-14	
-9,66817			-0,04535	0,090479	-0,00026	0,006832	-0,00287	0,013605	-4,4E-14	
0,441185	Chi-Sq	224,3507	-0,0122	-0,00026	0,000211	-0,0001	-7,3E-05	-0,00215	-3,8E-12	
0,056263	df	5	-0,07224	0,006832	-0,0001	0,127251	-0,00646	-0,00079	-4,9E-14	
2,805109	p-value	1,74E-46	-0,01699	-0,00287	-7,3E-05	-0,00646	0,008027	-0,00014	-2E-13	
0,214558	alpha	0,05	-0,15407	0,013605	-0,00215	-0,00079	-0,00014	0,320598	-5,1E-14	
1,406931	sig	yes								
	R-Sq (L)	0,411781		<i>coeff b</i>	<i>s.e.</i>	<i>Wald</i>	<i>p-value</i>	<i>exp(b)</i>	<i>lower</i>	<i>upper</i>
	R-Sq (CS)	0,198661	Intercept	-9,66817	1,075316	80,83814	2,45E-19	6,33E-05		
	R-Sq (N)	0,477562	Pol	0,441185	0,300797	2,151269	0,142452	1,554549	0,862115	2,803131
	AIC	332,4796	Vozrast	0,056263	0,014538	14,97794	0,000109	1,057876	1,028159	1,088453
	BIC	362,0036	Komplikac	2,805109	0,356722	61,83562	3,73E-15	16,52888	8,214871	33,25723
			Skor	0,214558	0,089595	5,734827	0,016632	1,239314	1,039723	1,477219
			Komorbid	1,406931	0,566214	6,17426	0,012962	4,083406	1,346048	12,38753

**Figure 1: Results from logistic regression**  
 Source: Author's calculations

**Probability of death**

The estimate of the probability of a certain event allows for an evaluation of the risk associated to the observed event, such as, the risk of death by a disease, the risk of getting cancer or heart attack, etc. The advantage of the calculated death odds from the previous chapter is the fact that the probabilities of death can be easily withdrawn from them by using the template  $odds/(1+odds)$  described in formula (6). So, for example, the probability of a deceased person by COVID-19 to be male is 0,61 or 1,55/(1+1,55). That means that 61 out of 100 deceased people are expected to be male. Furthermore, the people with complications have the highest expected probability to die from COVID, which is even 0,94 or 16,5/(1+16,53), while the expected probability of the deceased patient to have comorbidity is 0,82, that is 4,08/(1+4,08). This means that 94 out of 100 deceased patients would have

complications, while 82 would be people with comorbidity. The expected probability of a control group patient to die from the primary disease is estimated to be 0,0000632 or 0,0000633/(1+0,0000633), which means that 633 patients out of 10.000.000 would die. On the other hand, male patients at the age of 50 with comorbidity, complications and score 5 have expected death probability of 0,26 or 0,367/(1+0,357), which means that 26 out of 100 infected patients with these characteristics are expected to die.

**Significance of the Model (model fit)**

The model significance assessment is a useful tool for evaluation of the possibility of the model to fit into the whole population, but also it serves as a good approach to understand how good the chosen model reflects what is observed in the sample data. In fact, if the model with its determinants explains the observed



event better compared to the percentage of the realized outcome or event based on the basic model without determinants, than it is said that the model is statistically significant. The significance of a model is determined with the help of chi-square statistics which is a test done by comparing a null model without predictors (the percentage of people with the observed event) and the chosen model with predictors [18]. If the chi-square statistics is significant, that means that *the chosen model with predictors, expresses the probabilities of the observed event better than the null model without predictors*. Most software packages automatically derive the described calculations related to the chi-square statistics. In our case, the results are displayed in picture No.3, where it can be seen that our model is statistically significant, with chi-square value of 224,3507, degrees of freedom 5 and p-value lower than 0.001 (1,74E-46).

### The Assumptions of the Model

In order for any model to be statistically valid it needs to meet the assumptions or the criteria on which it is based. The model must meet all the assumptions, in order to be considered as generalized and unbiased out of the sample. The binary logistic regression is based on 3 basic assumptions: a) The observations in the statistical sample must be independent; b) Perfect multicollinearity cannot exist among the independent variables or predictors; and c) The continuous explanatory variables must be linearly related to the log-transformed values of the odds of the observed outcome (linearity) [19]. In addition we will add the assumption for the sample size, according to which the number of observations must be sufficient so that relevant and generalizable conclusions can be drawn based on the statistical model.

#### Linearity

According to this assumption, there must be a linear relation between the explanatory variables and the log-transformed values of death odds. Since part of our model is made of binary and ordinal (categorical) explanatory variables, the postulates of the linearity assumption remain valid only for the continuous variable. That means that in this particular case, it is important to define linear relation between the predictor *age* and the log-transformed values of death odds from the sample or *log(odds)*. One possible approach to detect the existence of described relation is through visualization with scatter plots. For this purpose, a scatter plot that explains the relation between the log-transformed values and the years of age was created in *Excel*. This is presented in graph No.1, where the linear relation between the two variables is clearly visible. Namely, as the age increases so the *log(odds)* values increase for the death outcome. Visually, the linear relation can be seen from the red line, while the corresponding linear relation is displayed

mathematically, through formula automatically created by *Excel*, in the lower right corner.

#### Independency

This assumption means that the observations in the sample should be mutually independent. Specifically, that means that the data should not be repeated (double or triple measurements for the same individual) and it should not originate from related individuals in any way. The latter could happen if there is grouping (people from a nursing home, kids from the same school or kindergarten) or if there are many members from the same family. The problem with the grouping occurs when people who come from the same family, school or nursing home, would probably share similar characteristics which can limit the variability of data and thus, could generate biased results [20]. In order to avoid violation of this assumption, the researcher needs to establish control over the whole process of data collection, that is, to be aware and to know in what way and from whom the data has been collected. Nevertheless, the data for our research was collected through a questionnaire and phone calls, and during the whole process of data collection there was no close proximity between patients on any grounds.

#### Multicollinearity

According to this assumption, there should not be any relation between the independent or explanatory variables. Multicollinearity appears when two or more independent variables mutually correlate between each other, so that they cannot provide unique or independent information within the regression model. If the degree of correlation between the observed variables is too high, it can create problems with the interpretation and fitting of the model [21]. Multicollinearity can be detected with the so-called variance inflation factor test (or VIF test), which measures the correlation and the intensity of correlation between the explanatory variables in the regression model. This test can be created by constructing a regression model only between the independent explanatory variables without the presence of dependent variable from the source model. For example, one of the independent variables is taken as a dependent variable and the relationship between that independent variable and the other independent variables is examined. Afterwards, this process is repeated for all independent variables with rotation, which means that if for example we have 5 explanatory variables such as in our model, as many regressions need to be made. VIF, is calculated for every independent variable according to the template (8):

$$VIF = \frac{1}{(1-R_{square})} \quad (8)$$

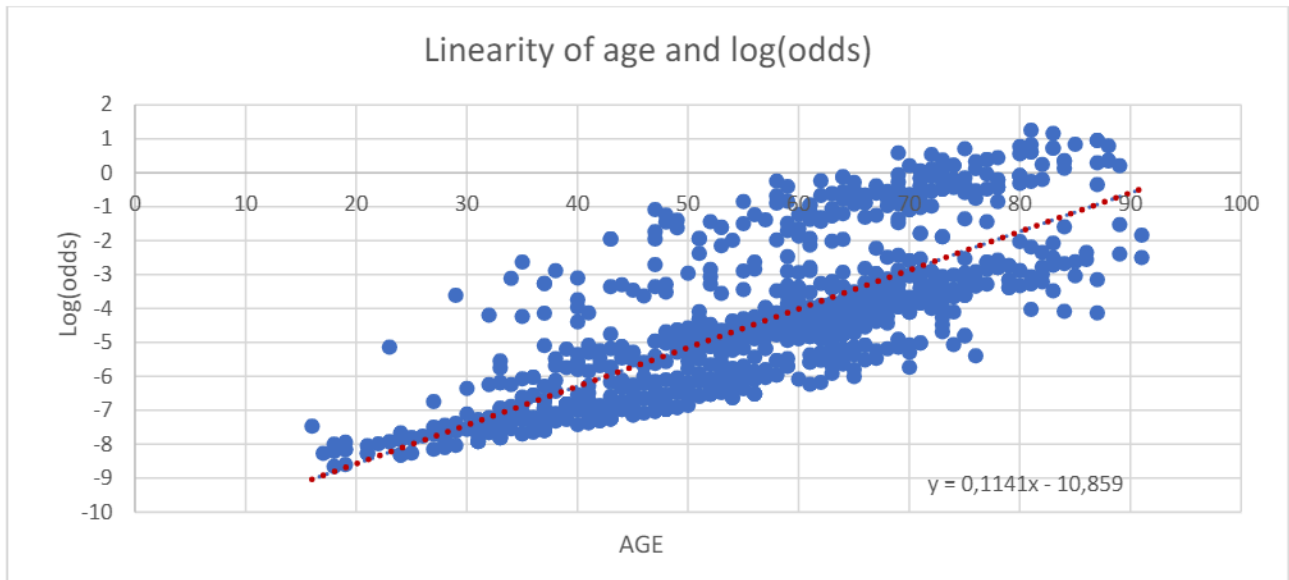
whose minimal value is 1, while the maximum is unlimited. If a value of 1 is obtained, that means that





there is no correlation between the examined explanatory variable and the other explanatory variables of the model. The value between 1 and 5 indicates average correlation between the variables, while the value higher than 5 indicates a strong correlation between the given independent variables and in such case, the regression coefficients and the p-value most probably won't be valid [22]. For this purpose, we have constructed 5 regressions within the framework of our research, one each for every

independent variable. According to the results, the VIF values are: 1,016 for *gender*, 1,233 for *complications*, 1,683 for *age*, 1,378 for *score* and 1,519 for the *comorbidity* variable. Also, we must mention that the average degree of correlation expressed through the VIF values in interval from 1 to 5, is generally not considered as a problem from aspect of the statistical analysis with logistic regression, and from this perspective, the chosen model satisfies the criteria of multicollinearity.



**Graph No.1. Scatter Plot of the Relation Between the Log-Transformed Values of Death Odds and Years of Age**  
 Source: Author's work;

Considering the fact that in medical research, the VIF value has generalized form, so, in such cases the use of the metric GVIF - generalized variance inflation factor is suggested. If the value of VIF is known, then it can be calculated as follows (9):

$$GVIF = VIF^{\left[\frac{1}{2 \cdot df}\right]} \tag{9}$$

where *df* are the degrees of freedom, or the number of independent variables in the model. Similarly, every value of GVIF which is higher than 2 is considered to be relating to a problem with the multicollinearity of the given model [23]. In our model the GVIF values for all explanatory variable are within the acceptable criteria (GVIF<2)

**Sample size**

Another critical assumption that must be met is the one related to the sample size because the strength of the predictive power of the statistical model depends mostly on the sample size or the number of observations. Namely, using a small sample for the development of prediction models can result in incorrect regression coefficients, which on the other hand can provide false and unstable predictions of the

outcome and generate weak prediction performances of the model [24]. There are many methodological approaches which have been developed and suggested in literature, aiming to define the minimal number of observations in a given medical study. However, in our research we will only use two criteria, out of which one is more general while the other one is more specific and frequently used within the medical and clinical trials. In order to evaluate "the effect of the sample size", these two approaches are explained in detail.

**General approach**

The general approach refers to the rule of *minimum 10 cases with the least frequent outcome* for each explanatory variable [25]. Minimum number of observations in the sample, with this method, can be obtained through the from (10):

$$N = \frac{10 \cdot \text{number of explanatory variables}}{\text{probability of least frequent outcome}} \tag{10}$$

So, for example, the least frequent outcome in our case is the outcome marked with 1 (death by COVID-19) while the more frequent outcome is the survival marked by 0. The probability of death outcome in our sample of 1013 COVID-19 patients is 0.076, while the



number of independent variables is 5. If these numbers are applied to the above form, the minimum number of patients required to meet the rule of the sample size is 658 patients. It is obvious that according to the rule of 10 cases with the least frequent outcome, our sample satisfies the criteria of the effect size.

#### **Specific approach**

The specific methodology approach, which is preferred by the medical studies based on logistic regression refers to the criterium *events per variable - EPV*, defined as ratio between the number of events or observations, relative to the number of degrees of freedom (number of independent variables) [26]. The existing form simplifies the complicated definition of the EPV criterium through importing *the rule of 50 events per variable*. Thus, according to this form, the minimum number of observations per study should be as follows (11):

$$N = 100 + 50 * \text{number of independent variables} \quad (11)$$

Incorporated into our example, this rule suggests that minimum 350 patients are needed in order to satisfy the assumption of the sample size. However, some authors suggest minimum number of 500 participants for the observation studies which involve logistic regression in their analysis, in order to generate parameters which would fit representatively into the targeted population [27].

#### **Discussion**

Scientific researches show that the total associated risk of COVID-19 is directly correlated to the gender, age and existence of secondary disease or comorbidity. The higher genetic expression of ACE2 receptor, especially in lung tissue [28] and most probably in the vascular epithelium of men, which is a gateway of the virus infection, is given as a possible reason for the higher mortality rate in male population, resulting in more furious development of the disease. The hormonal status stands out as a second possible causal relation between the COVID-19 death outcome and the gender differences. Different studies argue that female hormones such as estrogen and progesterone can act biologically protective against the viral infections, while the testosterone acts in a completely opposite way [29]. According to one study, women in menopause which were taking estrogen therapy had lower mortality rates in COVID-19 than the women without an appropriate estrogen therapy. The same study delivered a similar but inconclusive result from the research of the oncology patients with anti-estrogen therapy given to inhibit breast cancer and who seemingly had higher mortality rate in relation to the reference group [30]. The characteristic of the immune system of men can be systemized as another possible

reason for higher mortality rates of COVID-19 in the male population. According to researches, there is a more immediate connection between the male immune system and the probability of onset of a cytokine storm, and thus more damage to organs and possibility of death. Namely, in the case of infection with SARS-CoV-2, male population tends to produce more intensive inflammatory response, caused by, among the rest, the presence of protein *cytokine IL-6*, which presence is higher in male organism than in female. As a result of the increased cytokine response in male, they have higher chances of developing so called severe acute respiratory syndrome -SARS, and thus end with an unfavorable outcome [31]. After all, there are indications that the lower levels of testosterone in men could improve the prognosis of COVID-19 [32]. Also, men compared to women have higher predisposition for hypertension and heart diseases, and more often are prone to practicing certain habits which are hazardous to health, such as smoking, alcohol or drugs consumption [33]. Our research confirmed the strong relationship between the increased mortality rate in men and COVID-19, but such relationship was still insignificant (OR 1.55 with CI 0.86 - 2.80). When it comes to the determinant age, the clinical researches do not succeed to give unified answer to the type of mechanism which predisposes the elderly to a bad prognosis of the primary disease. The low immune response [34], obesity and the decline of respiratory function [35], general weakness of the older organism (frailty) [36] and multimorbidity [37] are given as possible hypothesis. According to another research, the genetic variations which support the longevity in elderly are significantly related to lower risk of infection with SARS-CoV-2 and hospitalization. A relationship has been detected between the acceleration of biological age and the possibility of future incidence with severe COVID-19 infection. The analysis suggests that the mechanism related to the acceleration of the aging process, which is the transmembrane receptor protein signaling pathway - Notch, is the same that leads to higher risk of COVID-19. Specifically, the higher genetic expression of Notch1 and Notch 2 in blood (especially in B-lymphocytes) is related to the increased risk of critical illness with COVID-19 [38]. Other studies relate the higher incidence of COVID in elderly to the aging of the immune system. The authors argue that the immune system needs to achieve 4 crucial functions: *recognition, alert, destruction and cleaning* of the viral pathogen in order to achieve effective suppression and elimination of SARS-CoV-2. It is known that there is a tendency that each of these mechanisms is becoming dysfunctional and incrementally heterogenous with aging. Namely, with time, the immune system goes through subtle changes in two ways. One relates to the



progressive decline in immune functions for recognition, alert and cleaning and the type of mechanism that is involved in this is called immunosenescence. The second one, is the progressive erosion of the function for destruction developed by the chronic increase of the systematic inflammation called inflammaging, which results from the hyperactive but ineffective signaling component of the immune system. The abundance of data describing the pathology and molecular changes in COVID-19 patients refer to the immunosenescence and inflammaging as main drivers to the high mortality rates in elderly [39]. In the context of better understanding the higher mortality cases in elderly, the relationship between the cell telomeres (as indicators of biological age) and COVID-19 death risks was also studied. However, it is worth mentioning that the results from the meta-study, contrary to the expectations, did not provide a strong evidence that shorter telomeres and higher biological age actually cause severe COVID-19 form and concluded that shorter telomeres do not pose a higher death risk factor [40]. Within the framework of our research, the obtained relationship between the years of age and COVID-19 mortality is also significant and positive (OR 1,06 with CI 1,03 – 1,09). When it comes to the third factor, which is comorbidity, we must emphasize that in this research we haven't done stratification of the type of comorbidity in order to analyze their individual influence towards the final outcome of the disease, taking into consideration the number of comorbidities, limited access to data, and the limitations of methodological nature. Instead, the aim was to examine the impact of comorbidity as a factor in general i.e. whether the presence of any chronic condition influences the disease and the intensity to which it influences the same. During the processing of data, presence of different comorbidities was detected in a large number of patients, among which the most common were the following: hypertension, cardiovascular diseases, diabetes, different coagulopathy (disseminated intravascular coagulation), kidney diseases, lung diseases, (COPD and asthma) different malignities obesity, etc. When it comes to primary coagulopathy, the following must be taken into consideration: even though the most severe cases of SARS-CoV-2 infection cause abnormalities in the coagulation process similar to DIC it is still significantly different than its usual clinical presentation and does not meet the international criteria for DIC. Namely, the coagulopathy related to COVID-19 is more specific with emphasized localized and prothrombotic (rather than hemorrhagic) characteristics which appear as a result of direct infection and damage to the vascular and alveolar epithelium [41]. The results from our research provide a solid proof that the existence of any type of

comorbidity, simultaneously to the primary disease, significantly increases the chances of mortality in the infected patients (OR 4,08 with CI 1,34 – 12,38). Also, the development of secondary lung complications or more specifically, bronchopneumonia, pneumonia and pleural effusion could significantly worsen the chances of favorable outcome, i.e., could significantly increase the mortality risk (OR 16,53 with CI 8,21 – 33,25). Also, positive significant relationship is detected between the radiological score of the primary disease and the mortality risk by COVID-19 (OR 1,24 with CI 1,04 – 1,47).

#### Limitations and Further Research Recommendations

Experience shows that the course of the disease COVID-19 could be highly unpredictable and could also be influenced by numerous factors. Thus, the developed model in this research encompassed the dominant conventional factors such as gender, age and the presence of comorbidity, but at the same time it was also upgraded with some unconventional determinants such as the presence of complications and the primary disease score. When it comes to complications and comorbidities, no distinction and stratification was made by their type, but their presentation and existence along with the disease was the crucial point. Of course different comorbidity and complications have different influence on the intensity and course of disease and that is why we recommend that additional research, with stratified data according to the comorbidity and complications type, is conducted. Vaccination is another potential factor of influence, which, sadly, was not possible to be integrated in the existing model. Namely, the vaccination process was conducted during the epidemic and it was impossible to incorporate this factor due to two reasons: first was the lack of data and second was the immunization period needed to develop immune response, which requires additional time. Given the subtlety of the immunization effect on the course of the disease, it is natural that its impact over morbidity and mortality rates to be analyzed in a separate study by using another unique and special statistical model.

#### Conclusion

The mortality outcome of COVID-19 is not an exclusive consequence of the primary disease and most often it is determined and in correlation with some other conditions, such as comorbidities and secondary complications, but also it depends on the other standard factors of influence such as age and gender. The proposed statistical model, built on the basis of binary logistic regression, is a relevant method to discover and measure the potential relationships among the patient's individual characteristics and the final outcome of a certain medical condition, expressed in the form of odds and probabilities. Hence its advantage for prognosis and assessment of the individual risk from the medical condition, which allows



the doctors to better plan and manage hospital protocols and treatment of patients.

### References

- [1] Worldometer. Available at: [www.worldometers.info](http://www.worldometers.info)
- [2] Sceznik C. Covid-19 and sex: higher mortality of male patients – a Report, Healthcare in Europe.com 2023. Available at: <https://healthcare-in-europe.com/en/news/covid19-sex-mortality-male-patients.html>
- [3] Peckham H, de Grijter NM, Raine C, Radziszewska A, Ciurtin C, Wedderburn LR, Rosser EC, Webb K, Deakin CT. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ICU admission. *Nat Commun.* 2020 Dec 9;11(1):6317. doi: 10.1038/s41467-020-19741-6
- [4] Taslem Mourosi J, Anwar S, Hosen MJ. The sex and gender dimensions of COVID-19: A narrative review of the potential underlying factors. *Infect Genet Evol.* 2022 Sep;103:105338. doi: 10.1016/j.meegid.2022.105338
- [5] Taslem Mourosi J, Anwar S, Hosen MJ. The sex and gender dimensions of COVID-19: A narrative review of the potential underlying factors. *Infect Genet Evol.* 2022 Sep;103:105338. doi: 10.1016/j.meegid.2022.105338
- [6] Sceznik C. Covid-19 and sex: higher mortality of male patients – a Report, Healthcare in Europe.com 2023. Available at: <https://healthcare-in-europe.com/en/news/covid19-sex-mortality-male-patients.html>
- [7] Chislett B, Chen D, Perera ML, Chung E, Bolton D, Qu LG. 5-alpha reductase inhibitors use in prostatic disease and beyond. *Transl Androl Urol.* 2023 Mar 31;12(3):487-496. doi: 10.21037/tau-22-690
- [8] Lightfoot C. Differences in COVID-19 outcomes related to sex. *News medical*; 2022.
- [9] Azizmohammad Looha M, Rezaei-Tavirani M, Rostami-Nejad M, Janbazi S, Zarean E, Amini P, Masaebi F, Kazemi M, Vahedian-Azimi A, Mirmomeni G, Jamialahmadi T, Guest PC, Sahebkar A, Pourhoseingholi MA. Assessing sex differential in COVID-19 mortality rate by age and polymerase chain reaction test results: an Iranian multi-center study. *Expert Rev Anti Infect Ther.* 2022 Apr;20(4):631-641. doi: 10.1080/14787210.2022.2000860
- [10] Ho FK, Petermann-Rocha F, Gray SR, Jani BD, Katikireddi SV, Niedzwiedz CL, Foster H, Hastie CE, Mackay DF, Gill JMR, O'Donnell C, Welsh P, Mair F, Sattar N, Celis-Morales CA, Pell JP. Is older age associated with COVID-19 mortality in the absence of other risk factors? General population cohort study of 470,034 participants. *PLoS One.* 2020 Nov 5;15(11):e0241824. doi: 10.1371/journal.pone.0241824
- [11] Vasudeva R, Challa A, Al Rifai M, Polana T, Duran B, Vindhyaal M, Lewis EF. Prevalence of cardiovascular diseases in COVID-19 related mortality in the United States. *Prog Cardiovasc Dis.* 2022 Sep-Oct;74:122-126. doi: 10.1016/j.pcad.2022.09.002
- [12] Zhou X, Cheng Z, Luo L, Zhu Y, Lin W, Ming Z, Chen W, Hu Y. Incidence and impact of disseminated intravascular coagulation in COVID-19: a systematic review and meta-analysis. *Thromb Res.* 2021 May;201:23-29. doi: 10.1016/j.thromres.2021.02.010
- [13] Puicón-Suárez JB, Zeña-Ñañez S, Failoc-Rojas VE. Association between chronic kidney disease and mortality in patients with a confirmed COVID-19 diagnosis. *PeerJ.* 2022 Jun 14;10:e13437. doi: 10.7717/peerj.13437
- [14] Meza D, Khuder B, Bailey JI, Rosenberg SR, Kalhan R, Reyfman PA. Mortality from COVID-19 in Patients with COPD: A US Study in the N3C Data Enclave. *Int J Chron Obstruct Pulmon Dis.* 2021 Aug 13;16:2323-2326. doi: 10.2147/COPD.S318000
- [15] Alhmoud E, Barazi R, Saad M, Al Khiyami D, El Ajez R, Bakdach D, Omrani A, Ibrahim W, El Anany R, Al-Hail M. Characteristics and outcomes of adult patients with asthma presenting with COVID-19: A comparative cohort study. *Qatar Med J.* 2023 Aug 8;2023(3):15. doi: 10.5339/qmj.2023.15
- [16] Abuhelwa Z, Alsughayer A, Abuhelwa AY, Beran A, Sayeh W, Khokher W, Sajdeya O, Khuder S, Assaly R. In-Hospital Mortality and Morbidity in Cancer Patients with COVID-19: A Nationwide Analysis from the United States. *Cancers (Basel).* 2022 Dec 30;15(1):222. doi: 10.3390/cancers15010222
- [17] Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Logistic regression. *Perspect Clin Res.* 2017 Jul-Sep;8(3):148-151. doi: 10.4103/picr.PICR\_87\_17
- [18] Harris JK. Primer on binary logistic regression. *Fam Med Community Health.* 2021 Dec;9(Suppl 1):e001290. doi: 10.1136/fmch-2021-001290
- [19] Harris JK. Primer on binary logistic regression. *Fam Med Community Health.* 2021 Dec;9(Suppl 1):e001290. doi: 10.1136/fmch-2021-001290
- [20] Harris JK. Primer on binary logistic regression. *Fam Med Community Health.* 2021 Dec;9(Suppl 1):e001290. doi: 10.1136/fmch-2021-001290
- [21] Bobbitt Z. The 6 assumptions of logistic regression; 2020. Available at: <https://rpubs.com/gkuleck/931252>



- [22] Bobbitt Z. The 6 assumptions of logistic regression; 2020. Available at: <https://rpubs.com/gkuleck/931252>
- [23] Harris JK. Primer on binary logistic regression. *Fam Med Community Health*. 2021 Dec;9(Suppl 1):e001290. doi: 10.1136/fmch-2021-001290
- [24] Dhiman P, Ma J, Qi C, Bullock G, Sergeant JC, Riley RD et al. Sample size requirements are not being considered in studies developing prediction models for binary outcomes: a systematic review. *BMC Medical Research Methodology*. 2023 Aug 19;23(1):188. doi: 10.1186/s12874-023-02008-1
- [25] Bobbitt Z. The 6 assumptions of logistic regression; 2020. Available at: <https://rpubs.com/gkuleck/931252>
- [26] van Smeden M, Moons KG, de Groot JA, Collins GS, Altman DG, Eijkemans MJ, Reitsma JB. Sample size for binary logistic prediction models: Beyond events per variable criteria. *Stat Methods Med Res*. 2019 Aug;28(8):2455-2474. doi: 10.1177/0962280218784726
- [27] Bujang MA, Sa'at N, Sidik TMITAB, Joo LC. Sample Size Guidelines for Logistic Regression from Observational Studies with Large Population: Emphasis on the Accuracy Between Statistics and Parameters Based on Real Life Clinical Data. *Malays J Med Sci*. 2018 Jul;25(4):122-130. doi: 10.21315/mjms2018.25.4.12
- [28] Taslem Mourosoi J, Anwar S, Hosen MJ. The sex and gender dimensions of COVID-19: A narrative review of the potential underlying factors. *Infect Genet Evol*. 2022 Sep;103:105338. doi: 10.1016/j.meegid.2022.105338
- [29] Taslem Mourosoi J, Anwar S, Hosen MJ. The sex and gender dimensions of COVID-19: A narrative review of the potential underlying factors. *Infect Genet Evol*. 2022 Sep;103:105338. doi: 10.1016/j.meegid.2022.105338
- [30] Elenis E, Kallner HK, Karalexi MA, Hägg D, Linder M, Fall K, Papadopoulos FC, Skalkidou A. Estrogen-modulating treatment among mid-life women and COVID-19 morbidity and mortality: a multiregister nationwide matched cohort study in Sweden. *BMC Med*. 2024 Feb 27;22(1):84. doi: 10.1186/s12916-024-03297-z
- [31] Sceznik C. Covid-19 and sex: higher mortality of male patients. *a Report, Healthcare in Europe.com* 2023. Available at: <https://healthcare-in-europe.com/en/news/covid19-sex-mortality-male-patients.html>
- [32] Chislett B, Chen D, Perera ML, Chung E, Bolton D, Qu LG. 5-alpha reductase inhibitors use in prostatic disease and beyond. *Transl Androl Urol*. 2023 Mar 31;12(3):487-496. doi: 10.21037/tau-22-690
- [33] Lightfoot C. Differences in COVID-19 outcomes related to sex. *News medical*; 2022.
- [34] Nikolich-Zugich J, Knox KS, Rios CT, Natt B, Bhattacharya D, Fain MJ. SARS-CoV-2 and COVID-19 in older adults: what we may expect regarding pathogenesis, immune responses, and outcomes. *Geroscience*. 2020 Apr;42(2):505-514. doi: 10.1007/s11357-020-00186-0
- [35] Sattar N, McInnes IB, McMurray JJV. Obesity Is a Risk Factor for Severe COVID-19 Infection: Multiple Potential Mechanisms. *Circulation*. 2020 Jul 7;142(1):4-6. doi: 10.1161/CIRCULATIONAHA.120.047659
- [36] Montero-Odasso M, Hogan DB, Lam R, Madden K, MacKnight C, Molnar F, Rockwood K. Age Alone is not Adequate to Determine Health-care Resource Allocation During the COVID-19 Pandemic. *Can Geriatr J*. 2020 Mar 1;23(1):152-154. doi: 10.5770/cgj.23.452
- [37] Ho FK, Petermann-Rocha F, Gray SR, Jani BD, Katikireddi SV, Niedzwiedz CL, Foster H, Hastie CE, Mackay DF, Gill JMR, O'Donnell C, Welsh P, Mair F, Sattar N, Celis-Morales CA, Pell JP. Is older age associated with COVID-19 mortality in the absence of other risk factors? General population cohort study of 470,034 participants. *PLoS One*. 2020 Nov 5;15(11):e0241824. doi: 10.1371/journal.pone.0241824
- [38] Ying K, Zhai R, Pyrkov TV, Shindyapina AV, Mariotti M, Fedichev PO, Shen X, Gladyshev VN. Genetic and phenotypic analysis of the causal relationship between aging and COVID-19. *Commun Med (Lond)*. 2021 Oct 5;1:35. doi: 10.1038/s43856-021-00033-z
- [39] Mueller AL, McNamara MS, Sinclair DA. Why does COVID-19 disproportionately affect older people? *Aging (Albany NY)*. 2020 May 29;12(10):9959-9981. doi: 10.18632/aging.103344
- [40] Haridoss M, Ayyasamy L, Bagepally BS. Is COVID-19 severity associated with telomere length? A systematic review and meta-analysis. *Virus Genes*. 2023 Aug;59(4):489-498. doi: 10.1007/s11262-023-02010-1
- [41] Levi M, Iba T. COVID-19 coagulopathy: is it disseminated intravascular coagulation? *Intern Emerg Med*. 2021 Mar;16(2):309-312. doi: 10.1007/s11739-020-02601-y

