Investigating the Male and Older People Susceptibility to Death from (COVID-19) Using Statistical Models

Rabia Emhamed Al Mamlook*, Zakaria Hashi†, Tiba Zaki Abdulhameed‡ and Hanin Fawzi Bzizi‡

1Department of Industrial Engineering and Engineering Management, Western Michigan University, Kalamazoo, MI, United States
2Department of Statistics, Western Michigan University, MI, United States
3Department of Computer Science Al-Nahrain University, Baghdad, Iraq
4Department of Biomedical Science, Western Michigan University, MI, United States

Abstract

Introduction: Coronavirus disease 2019 (COVID-19) is one of the serious infectious diseases that is caused by a specific virus called SARS-CoV-2. The rapid spread of COVID-19 raises serious concerns about the globally growing death rate. Currently, cases are doubled in one week around the world. Recorded data shows that COVID-19 does not infect all patients equally. This opportunistic virus can affect people of any age and gender. Information about the reason for high mortality in the age group 60 and older is limited. The gender differences among all deceased are poorly known. To understand more about COVID-19, this study aims to examine the different age groups among the death and focuses on comparing genders between males and females.

Method: Statistical analysis including Pearson’s Chi-squared ($\chi^2$) and binary logistic regression was conducted based on existing data to examine factors relating to death, such as age and gender. Adjusted odds ratios (ORs) with 95% confidence intervals (CIs) were calculated for death.

Results: The results show that males were 2.51 more likely to die of coronavirus COVID-19 than females. Moreover, the study found a significant increase in death for patients age 60 and older compared to patients age less than 40. Thus, males of 80+ age were found to be highly associated with death.

Conclusions: Older people and male are more susceptible to death from COVID-19, we should pay more attention to the elderly people and male with COVID-19. This imposes providing careful health care for this population.

Keywords: COVID-19 • Death • Age • Odds Rate • Logistic Regression • Older People • Male

Introduction

COVID-19 is an infectious disease caused by a specific virus called SARS-CoV-2. The COVID-19 outbreak was first reported in Wuhan, China in late December 2019 [1-3] and it has spread quickly in many other countries, including Europe and the United States causing it to become a global health emergency [3]. As of March 24, 2020, according to the World Health Organization (WHO) and figures from state government leaders and health officials, over 523,163 cases have been confirmed and more than 23,639 have died of the virus since the start of the spread [3]. Figure 1 illustrates how the number of deaths is increased weekly for all countries. A study based on 1099 Covid-19 patients from 552 hospitals in 30 provinces in mainland China through January 20, 2020, showed patients often presented without fever, and many did not have abnormal radiologic findings [4].

This has been a huge challenge to distinguish Covid-19 patients from healthy ones. Many cases of COVID-19 are mild and can recover quickly, but some cases can be severe and deadly, with the highest mortality rate of around 3.4% [3,5]. Scientists work around the clock to test medications for healthy ones. Many cases of COVID-19 are mild and can recover quickly, without fever, and many did not have abnormal radiologic findings [4].

*Address for Correspondence: Rabia Almamlook, Department of Industrial and Entrepreneurial Engineering & Engineering Management, Western Michigan University, 1903 W. Michigan Ave., Kalamazoo, MI 49008-5336, United States, Tel: +12693830404, Email: rabiaemhamedlm.almamlook@wmich.edu

Copyright: © 2020 Almamlook R, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received 17 June 2020; Accepted 30 July 2020; Published 06 August 2020

was observed with 19 lopinavir-ritonavir treatment beyond standard care [6]. Another study found that age was mortality risk factors. This study aimed to identify risk factors for mortality in elderly patients with COVID-19 [7]. This study used age as risk factor for predicting mortality in elderly patients with COVID-19. Moreover, elderly people has been one of the largest problems in several countries that there are many old people who will face the risk of infected with COVID-19, which impressive a heavy problem on the health care systems in the world [8,9].

This opportunistic virus can affect all people of any age or gender. Early reports of the outbreak in China suggested that males were especially at risk. Therefore, a study of 99 patients at a hospital in Wuhan, where the virus originated, found that males made up two-thirds of patients. It showed a strong gender breakdown of deaths, which were 84% male [10]. In a recent study published in the Lancet, found that 80% of the deaths were in males and just 20% were in females [11]. Previous studies of COVID-19 were based on information from the general population, limited data are available for patients with COVID-19. Another study found that age was mortality risk factors. This study aimed to identify risk factors for mortality in elderly patients with COVID-19. There is limited research on COVID-19 by age and gender. The primary purpose of the study is to examine death rates by gender, age, and both age and gender using various statistical analyses (Figure 1).

Therefore, the main purpose of this study is to conduct a statistical analysis on factors such as age group and gender that may result in death from COVID-19 and associate those among them that have a more pronounced impact on comprehensive analysis. In this study, we analyzed using binary logistic regressions by assuming the death is independent among group age and gender. The remainder of the paper is organized as follows. Section 2 discusses the data and Methodology. Section 3 describes the results and discussion. Finally, Section 4 concludes the paper and presents directions for 48 future researches.
Data and Methodology

Description of the Date

The analyses in this study are based on the COVID-19 dataset. The data was obtained from the Kaggle [12] that allows data analysts to compete with each other’s to solve real and complex data knowledge problems. All patients admitted to the hospital and diagnosed with COVID-19 from January 2020 to March 2020 were included in this study. The data contains 1086 record. Variables that were considered in the models included: gender, age, and death. Additionally, patients 57 were classified into 9 groups according to age including 0-9; 10-19; 20-29; 30- 58 39; 40-49; 50-59; 60-69; 70-79 and 80+ years old. Age groups less than 39 are 59 excluded from this study due to no death cases.

Statistical methods

Descriptive statistics were used to identify the potential factors that had a statistically significant influence on the likelihood of death. Chi-square test ($\chi^2$) of homogeneity was used to test the relationship between the potential predictors (age and gender) on the outcome (death). The test is a non-parametric test with no assumed distribution. It has been used broadly as it does not execute conditions in the data, such as equality of variance or residual homoscedasticity. Therefore, to further understand this relationship, the logistic regression model was developed to identify the degree of significance of each independent predictor. The null hypothesis the authors wish to reject in this test is that there is no significant relationship between two variables. For this study, we rejected the null hypothesis if the p-value was less than 0.05. Chi-square statistics can be computed as shown in the equation below:

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$  \hspace{1cm} (1),

where $O_{ij}$ is the observed frequency and $E_{ij}$ is the expected frequency across row $i$ and column $j$ of the contingency. The calculated $\chi^2$ is associated with the critical value found from Chi-square distribution. The definite degrees of freedom (df) for the critical value can be subtracted as $(c-1)*(r-1)$ where $c$ signifies the number of columns and $r$ represents the number of rows.

Logistic regression Model

Binary logistic regression was used because it is a preferred method when the response variable is dummy variable [13]. The logistic regression estimated the probability of a patient to choose either death or recover to, given a set of explanatory or predictor variables. Mathematical formulation of logistic regression can be presented as shown in the equation below:

$$\ln\left(\frac{p}{1-p}\right) = \sum_{i=0}^{c} \beta_i x_i$$  \hspace{1cm} (2),

whereby, $p$ is the probability of a person death or recover and $x_i$ is the explanatory variable of interest with its corresponding coefficient $\beta_i$.

Here, $\beta_i$ is the coefficient of the predictor or input variable used in a regression equation [14]. The logit is the natural logarithm of the odds ratio (OR), ranging from 0 to positive infinity. When an independent variable $x_i$ increases by one unit, with all other factors remaining constant, the odds increase by a factor $exp$ which is called the odds ratio (OR) given the odds ratio (OR) range from 0 to positive infinity. It indicates the relative amount by which the odds of the outcome (death) increase (OR>1) or decrease (0<OR<1) when the value of the corresponding independent variables increases by one unit.

Odds ratio

Odds ratio (OR) have become commonly used in medical studied [15]. OR is used to compare the relative odds of the occurrence of the outcome of interest (death) given exposure to the variable of interest (age and gender). 95% confidence intervals (CIs) are calculated to estimate the precision of the odds ratio. Mathematical formulation of odds ratio can be presented as shown in the equation below:

$$\frac{p}{1-p} = e^{\sum_{i=0}^{c} \beta_i x_i}$$  \hspace{1cm} (3),

Whereby, $p$ is the probability of a person death or recover and $x_i$ is the explanatory variable of interest with its corresponding coefficient $\beta_i$. The term $\frac{p}{1-p}$ are called odds ratio of event.

- OR = 1 Indicate does not affect odds of outcome
- OR > 1 Indicate a positive relationship with higher odds of outcome (event likely to occur)
- OR < 1 Indicate a negative relationship with less odds of outcome (event less likely to occur).

Results and Discussion

Statistical analysis

During the study period, 1086 patients with COVID-19 were admitted to the hospital. Before analysis, 260 of missing observations were excluded. A total of 825 patients with COVID-19 confirmed influenza during the January–March were included in the study. Total male cases are 476 (56%), while total female cases are 349 (42%). Table 1 shows the statistical description of the data. The recovery rate, death rate, and mean of age described as counts and percentages Figure 2.

The overall death rate is 7% (n=58) in this study. The death rate in males is 9% (n=44), while it is only 4% (n=14) in females. This makes the rate is 5% higher in males than in the female. However, if we look at the total death cases alone, Males are 23 times higher than females. Elderly citizens over the age of 70 represent more than 50% of the deaths combined, likely reflecting the presence of other diseases, a weaker immune system, or
simply worse overall health.

To visualize the data, we use one of the most common statistical charts is histograms visualize to understand frequency distribution of a given set of continuous data. An illustration of the age and gender distribution is shown in Figure 2.

It shows more males than females across all age groups. Figure 3 describes the relationship between age and death. The graph indicates that the death rate increases with age. The risk of death is higher in people age 60 or older. Figure 4 shows that more males died from the virus than females. However, the impact of gender on susceptibility to COVID-19 is less strong than the age effect.

Modal Choice Predictors:

A chi-square test of independence was performed to examine the relationship between death and related factors such as age and gender. Table 2 shows the relationship between gender and death was significant with \(^{\chi^2}\) test = 8.434 and (p = 0.0036). Similarly, it shows that the relationship between age and death was also significant with \(^{\chi^2}\) test = 60.853. A significant association was found between death, age group, and gender using a \(^{\chi^2}\) test (P less than 0.05). The result indicates that age and gender enable the model to make better predictions.

To further analyze the data, a binary logistic regression model was used. A dummy variable for a response variable was created with 1 assigned to death and 0 for recover. The results of the logistic regression examining the association between death and all variables are shown in Table 3. The logistic regression model can be presented as shown in the equation below:

\[
p(\text{log} \gamma) = 4.5549 + 0.6773 \cdot \text{Age}_{50-59} + 1.7299 \cdot \text{Age}_{60-69} + 2.1 \cdot \text{Age}_{70-79} + 3.4259 \cdot \text{Age}_{80+} + 0.9209 \cdot \text{Gender}_{Male} \quad (4)
\]

According to the model, the log odds for dying the coronavirus was positively correlated to age but the correlation was statistically significant to age group 60 and older (p < 0.05), in another words as age increases likelihood for dying the virus increases. Both age and gender were significantly associated with death with \((p<0.05)\). According to the model, the coefficients were positively correlated to death, meaning that when age increases the rate of death increases as well.

Table 4 presents the results with odds ratios, coefficient values, and the p-value. The estimated coefficient for the intercept shown in Table 4 is the exponential of the odds ratio for age groups (40-49) in Table 4, in other words, the odds in intercept in Table 4 equals to \((0.0125)\) that is exp \((-4.55)\). The result in Table 4 indicates that there is an increase of 1.25% (98.75%) in the odds of death for every year increase in age groups (40-49) of the patients. If we look at the counts of the death cases in Table 3. we can see that the number of females who died in age groups 40-49 was one, and only two for males. Thus, COVID-19 is less likely to cause death in patients age less than 49. For everyone unit increases in the age group (50-59), the chances of death increase by 96.8%. Furthermore, the odds of COVID-19 deaths in the age group (60-68) are about 5.6 more likely than those in the

---

**Table 1. Descriptive Statistic.**

<table>
<thead>
<tr>
<th>Gender (n = 825)</th>
<th>Men (n = 478)</th>
<th>Women (n = 349)</th>
</tr>
</thead>
<tbody>
<tr>
<td>alive (n = 432)</td>
<td>% (0.91)</td>
<td>% (0.99)</td>
</tr>
<tr>
<td>Death (n = 44)</td>
<td>% (0.09)</td>
<td>% (0.04)</td>
</tr>
<tr>
<td>alive (n = 335)</td>
<td>% (0.96)</td>
<td>% (0.96)</td>
</tr>
<tr>
<td>Death (n = 14)</td>
<td>% (0.04)</td>
<td>% (0.04)</td>
</tr>
<tr>
<td>0-9 9</td>
<td>0.56 ± 2.8</td>
<td>5.0 0.42 ± 2.8</td>
</tr>
<tr>
<td>10-19 6</td>
<td>0.16 ± 2.8</td>
<td>6.0 0.16 ± 2.8</td>
</tr>
<tr>
<td>20-29 48</td>
<td>0.25 ± 2.1</td>
<td>44.0 0.25 ± 2.6</td>
</tr>
<tr>
<td>30-39 83</td>
<td>0.34 ± 2.6</td>
<td>60.0 0.34 ± 2.7</td>
</tr>
<tr>
<td>40-49 77</td>
<td>0.44 ± 2.6</td>
<td>46.1 0.44 ± 2.7</td>
</tr>
<tr>
<td>50-59 98</td>
<td>0.64 ± 2.25</td>
<td>76.1 0.55 ± 2.4</td>
</tr>
<tr>
<td>60-68 62</td>
<td>0.64 ± 2.3</td>
<td>56.4 0.63 ± 12.7</td>
</tr>
<tr>
<td>70-79 35</td>
<td>0.73 ± 2.2</td>
<td>32.3 0.73 ± 2.3</td>
</tr>
<tr>
<td>80+ 13</td>
<td>0.84 ± 3.2</td>
<td>10.5 0.86 ± 4.7</td>
</tr>
</tbody>
</table>

**Table 2. Pearson’s Chi-squared (\(^{\chi^2}\)) distribution of death by gender and age group.**

<table>
<thead>
<tr>
<th>(^{\chi^2}) - value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death-Gender 8.434</td>
<td>0.00368</td>
</tr>
<tr>
<td>Death-Age Group 60.853</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

**Table 3. Result of the Logistic Regression Model-Multivariate.**

<table>
<thead>
<tr>
<th>Model</th>
<th>Coefficients</th>
<th>Standard Error</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept -4.5549</td>
<td>0.6455</td>
<td>0.35623</td>
<td>&lt;0.0001 ***</td>
</tr>
<tr>
<td>Age_50-59 0.6773</td>
<td>0.6888</td>
<td>0.00760 **</td>
<td></td>
</tr>
<tr>
<td>Age_60-69 1.7299</td>
<td>0.6480</td>
<td>0.00116 **</td>
<td></td>
</tr>
<tr>
<td>Age_70-79 2.1500</td>
<td>0.6816</td>
<td>0.00011 **</td>
<td></td>
</tr>
<tr>
<td>Age_80+ 3.4259</td>
<td>0.6751</td>
<td>&lt;0.0001 **</td>
<td></td>
</tr>
<tr>
<td>Gender_Male 0.9209</td>
<td>0.3392</td>
<td>0.00663 **</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4. Odds Ratio and Confidence interval (95 percent CI) to all significant.**

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Odds Ratio</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 0.0125</td>
<td>0.0028-0.0382</td>
<td>1.21e-12 **</td>
<td></td>
</tr>
<tr>
<td>Age Group_50-59 1.968</td>
<td>0.5550-9.1456</td>
<td>0.35623</td>
<td></td>
</tr>
<tr>
<td>Age Group_60-69 5.639</td>
<td>1.7940-24.8923</td>
<td>0.00760 **</td>
<td></td>
</tr>
<tr>
<td>Age Group_70-79 8.584</td>
<td>2.6364-38.5943</td>
<td>0.00116 **</td>
<td></td>
</tr>
<tr>
<td>Age Group_80+ 30.749</td>
<td>9.2355-141.4070</td>
<td>3.87e-07 **</td>
<td></td>
</tr>
<tr>
<td>Gender_men 2.511</td>
<td>1.322-5.0398</td>
<td>0.00663 **</td>
<td></td>
</tr>
</tbody>
</table>

The result in Table 4 indicates that there is an increase of 1.25% (98.75%) in the odds of death for every year increase in age groups (40-49) of the patients. If we look at the counts of the death cases in Table 3. we can see that the number of females who died in age groups 40-49 was one, and only two for males. Thus, COVID-19 is less likely to cause death in patients age less than 49. For everyone unit increases in the age group (50-59), the chances of death increase by 96.8%. Furthermore, the odds of COVID-19 deaths in the age group (60-68) are about 5.6 more likely than those in the
age group (40-49).

Similarly, the odds of deaths in age group (70-79) increases by 9 times more than the age group (40-49). Moreover, in this study patients in the age group, 80+ are at most risk as their odds of death are 30 times greater than those in the age group (40-49).The 95% confidence interval in the age group 80+, shown in Table 4, is wider than the confidence interval of other groups because of the small size of the sample in this group. Considering the gender factor, males death odds as compared to the females, where the odds ratio of males equals 2.5. Thus, the odds of death in males is 2.5 times greater than the odds of death in females. A major risk factor in both genders but stronger in males than in females. Therefore, elderly male citizens are more susceptible to coronavirus as compared with other groups.

In summary, the correlation was statistically significant to the age groups of 60 and older where (p<0.05). Thus, when age increases, the likelihood of dying from the virus increases. This study proves that the death risk of COVID-19 is at its higher ratio in males aged 60+. There are many reasons why males are more likely to die from COVID-19. The most likely explanation for smoking is a risk factor for COVID-19 is that males are more likely to smoke than females [16]. Older adults are also more likely to suffer from some complicating conditions, like heart disease, lung disease, and diabetes that put them at higher risk for severe outcomes from contracting COVID-19 [17]. Likewise, females generally having greater or more robust immune responses than males [18]. Research shows that women have stronger immune responses to coronaviruses.

However, many limitations should not be ignored. One limitation is the limited availability of complete data on which to conduct the analysis. Analysis of the data samples of this study reveals that the death was highly correlated with age and gender. So, while the analysis produced non-significance, it is anticipated that as more data becomes available, the models will yield more concrete findings. Some other important factors may affect COVID-19 mortality, such as smoking and complicating conditions, like heart disease, lung disease, and diabetes. Therefore, these issues should be examined in future studies. Regardless, understanding the relationships among death causal factors may shed light on those causal factors which have the potential to lead to know why men and older are more likely to die from COVID19. The larger data sample size is needed for further study to better understand the susceptible group to COVID-19.

Conclusion

This study investigates if older people and males are more at susceptible to the novel Coronavirus. In this regard, the results revealed that death probability was significantly related to age and gender. This indicated that males are more likely to die of COVID-19 than females by 2.5 times. Our findings also indicate an increase in the probability of death in patients aged 60 and older. The COVID-19 infection is generally susceptible with death in older people and male, we should pay more attention to the elderly people and male with COVID-19 infection. Further studies are needed to understand why mortality is more strongly associated with males than in females. This implies that further studies need to analyze the causes of having higher death risk in males to protect this population and reduce the overall risk factor. Further studies are needed to understand why mortality is more strongly associated with men than in women according to some factors such as smoking and complicating conditions.

References


