Statistical Similarity of Mortality and Recovery Ratios for Covid-19 Patients based on Gender and Age

Abbas Mahmoudabadi 1*

1 Director, Master Program in Industrial Engineering, MehrAstan University, Guilan, Iran.

Received 14 September 2021; Revised 11 November 2021; Accepted 18 November 2021; Published 01 December 2021

Abstract

Background: Studying the behavior of patients infected to Covid-19 is an essential issue for health authorities during the global pandemic, so the aim of this study is to investigate the statistical similarity between the recovery and mortality ratios based on the patients’ age and gender. To this purpose, the well-known statistical testing method of Kolmogorov-Smirnov has been utilized to investigate the similarity of distribution functions for mortality and recovery rates for patients infected to Covid-19. Results: Data for 1015 patients resulted in dead, recovery, and transferred have been collected and analyzed. The age is cross-classified by gender where the rates’ cumulative distribution functions are independently calculated and depicted for females and males. The results revealed there is no significant difference between the distribution functions of mortality and recovery rates by gender but there is by age. Conclusion: The research results would support the health authorities to manage the admission and discharging procedures of the Covid-19 patients where the hospitality services are traditionally provided differently by gender.

Keywords: Covid-19; Distribution Function; Statistical Similarity; Mortality and Recovery Rate.

1. Introduction

Covid-19, the latest mutation of Coronavirus, is a viral infection and a respiratory disease with rapid human-to-human transmission on air. It poses health risks to the patients with weak immune systems, heart or kidney diseases, and pregnant women as well [1, 2]. Patients infected to Covid-19 may also experience a wide variety of symptoms like fever, cough, and shortness of breath, and even gastrointestinal problems. For the elderly people, more suffering symptoms such as lethargy, weakness, fatigue, mood swings, and decreased concentration may also appear [3]; while the patients with underlying diseases are likely to be more severely suffered like cardiac arrhythmia, urinary output (anemia), seizures, loss of consciousness, bleeding, shock, and pulmonary edema [4].

During a pandemic outbreak, hospitals and healthcare systems’ managers are primarily involved in managing resources including bed, staff, and equipment to resolve the health problems [5]. One of the main concerns for healthcare authorities is the difference between females and males, in particular, where they need to receive different healthcare or treatment operations in separate places due to some restrictions such as religious considerations. Therefore, it is necessary to manage the nurses and healthcare facilities as well as the patients, because there are a limited number of equipped beds in the hospitals. Supporting healthcare authorities to gain a deeper understanding of similarities and dissimilarities between mortality and recovery rates by gender and age would be an important issue in this field. The above perceptive would support decision-makers to manage hospitalized operations for the patients who

*Corresponding author: mahmoudabadi@mehrastan.ac.ir

http://dx.doi.org/10.28991/HIJ-2021-02-04-05

This is an open access article under the CC-BY license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.
may receive exclusive healthcare based on their age and gender. In other words, perceiving the patients’ behaviors for predicting the mortality and recovery rates supports medical authorities to manage healthcare operations as well as nursing capabilities, so the research is to try to develop the above concern.

1.1. Distribution Similarity and Concept

Checking the similarity of distribution functions is one of the practical methods to examine the relevancy between dependent and independent variables. In this case, two or more data sets are compared based on the similarity of their distribution functions through utilizing statistical tests [6]. There are many measures which could be evaluated to check the similarity of two distribution functions [7], but they are dependent on which method is used for this purpose. The Kolmogorov-Smirnov test, known also as the KS test and primarily utilized in non-parametric hypothesis testing [8], is one of them that compares the behavior of two related samples where each record in one population and it is compared individually to the same observation in the other population. It is a goodness of fit and non-parametric test of the equality of continuous or discrete one-dimensional probability distributions to compare the statistical probabilities of two samples [9], so it is conventionally utilized to compare a real distribution sample with a reference probability distribution [10].

The principal concept behind the KS test is to investigate the distance between the cumulative distribution functions of two samples which represents the unlikelihood of two distribution shapes [11]. Therefore, it is utilized to assess the similarity between the expected and the experimental or observational distribution functions for checking the fitness of the experimental data to the expected distribution function [12]. Although this ability commonly supports data analyzers for testing normality, where the existing normality is necessary to perform analysing procedures [13], it can be also utilized in other distribution functions and existing similarities for two data sets [14].

1.2. Relevant Studies

Utilizing the statistical methods is repeatedly observed to examine the relevancy of variables assessed in health research fields [15]. In the last decades, many studies have been conducted for investigating the relationship between healthcare operations and their basic requirements. The health-based recorded data has been frequently analyzed utilizing data mining techniques and multiple linear regression methods to develop models and provide accurate estimations [16]. For example, in the Tsuyama Hospital, Japan, a study has been conducted for predicting the cost of public healthcare to manage hospitalized operations based on developing a linear regression, recorded observations, and showed the forecasting models are capable to predict the health care costs [17].

Studies on estimating the mortality and recovery ratios of patients or other medical measurements are also observed in the literature [18]. For example regression analysis has been widely utilized across the populations [19] and predicting the recovery rate has been studied repeatedly [20]. Studies have been moreover conducted on diseases or related symptoms like examining the prognosis and likelihood of heart disease with various symptoms, because heart disease kills one person every 40 seconds, according to the American Heart Association [21]. Comparing situations is another field of studies in this area whereas, in terms of long-term healthcare, the study of the healthcare system forecast and its impact on health costs through linear regression in Colombia showed that residence long-term healthcare is costly for insurers and patients. For example, the Johns Hopkins University conducted a study in 2020 to predict the prevalence of Covid-19 and determined the most contributing factors in disease outbreak in short term [22] as well as it predicts significant savings in the patients’ health costs according to the health policy decisions [23].

The unknown Coronavirus has being attracted more rapidly under research in recent months since it was getting to an outbreak since December 2019. Age is a very important contributing factor for all patients’ recovery and mortality ratio [24], as well as for the patients infected toCovid-19, where elderly is one of the risk factors of increasing the mortality rate of the patients [25]. All parameters and information related to Cvid-19 patients, including age, gender, symptoms, and underlying disease status should be more investigated to manage healthcare operations. Because the Iranian big cities had received many travelers from the other countries of UAE, Chian, Oman, and Iraq at the beginning of the outbreak, the studies on Coronavirus have been also conducted in the country at the time [26] meanwhile spreading the virus has not been stopped. Moreover, in the field of the incubation period [27], asymptomatic ratio [28], epidemiological parameters and epidemic predictions [29], risk transmission, and even estimating the number of confirmed persons infected by Coronavirus [30] have been dramatically attracted by healthcare researchers. Therefore, more studies are required to focus on the healthcare system in Iran where it is necessary to know more about the virus outbreak.

1.3. Contribution Statement

Following the above mentioned, the study has been conducted to investigate the similarities and dissimilarities between the mortality and recovery rates according to patients’ gender and age, in which the novelty behind the research work lies on the differences of distribution functions fitted for mortality and recovery rates based on the mentioned attributes of age and gender. In other words, studying the females and males mortality and recovery rates is necessary to know more about the virus outbreak.
2. Research Methodology

As stated in the previous section, the comparison of distribution similarities between mortality and recovery rates for females and males is now investigated for Covid-19 patients in different age groups and gender. The main stages of statistical analysis include description on data collection, defining hypotheses followed by utilizing the Kolmogorov-Smirnov throughout this section.

2.1. Case Study and Data Collection

To implement the research methodology, two types of variables should be clearly defined. The first type is the group of variables that specify personal characteristics including gender and age, whereas the second group comprises of resulting variables like mortality and recovery rates. Data for 1015 patients infected to Covid-19 were collected from February 18 to August 20, 2020, in the northern Iranian province of Guilan. Three designated hospitals, where patients had been separately under intensive care, were selected as the case study. Data have been collected for six months from the start time of the outbreak to August 2020 through the Health Information System abbreviated as HIS in the Islamic republic of Iran. They are composed of many recorded fields in which personal specifications of the age and gender and the type of clearance were available. They recorded 1015 patients of 427 women and 588 men in which 161 patients resulted in dead, 146 discharged as personal satisfaction, 603 recovered, and finally, 105 transferred to the other hospitals or homes. Personal satisfaction means the case that the patient is discharged according to the family or his/her request mainly for staying at home. Fortunately, HIS provides all the required fields that help authors to conduct the study. The descriptive stats and more details of the patients who have been categorized by gender and age have been tabulated in Table 1 that demonstrates an overall view of what collected and analyzed during the study. As shown, gender is categorized into the female and male, age into ten categories from zero to 100 years old stepping by ten years.

<table>
<thead>
<tr>
<th>Age Group/ Gender</th>
<th>Death</th>
<th>Personal Satisfaction</th>
<th>Recovered</th>
<th>Transferred</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>M</td>
<td>F+M</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>00-09</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10-19</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>20-29</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>30-39</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>40-49</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>50-59</td>
<td>11</td>
<td>14</td>
<td>25</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>60-69</td>
<td>13</td>
<td>16</td>
<td>29</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>70-79</td>
<td>17</td>
<td>27</td>
<td>44</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>80-89</td>
<td>19</td>
<td>17</td>
<td>36</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>90-100</td>
<td>4</td>
<td>5</td>
<td>9</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>90</td>
<td>161</td>
<td>63</td>
<td>83</td>
</tr>
</tbody>
</table>

F=Female; M=Male

2.2. Defining the Hypotheses

The first step of performing the test is to define its hypothesis. Since the Kolmogorov-Smirnov test is utilized to check the similarity or dissimilarity between females and males in terms of mortality and recovery rates, the null and competitive hypothesis are defined as follows whereas it is assumed that the patients who have been discharged based on their personal satisfaction, have been categorized as recovered patients.

- \( H_0 \): Mortality and recovery rates of females and males come from the same distribution functions.
- \( H_1 \): Mortality and recovery rates of females and males come from different distribution functions.

The hypotheses for all patients are defined as follows where the test is utilized to compare the rates for all patients.

- \( H_0 \): Mortality and recovery rates of all patients come from the same distribution functions.
- \( H_1 \): Mortality and recovery rates of all patients come from different distribution functions.
3. Results

The mortality and recovery rates have been calculated based on the data received from HIS and tabulated in Table 2. The domain of age is divided into ten categories from zero to 100, and the above-mentioned rates have been individually calculated for females and males. For example, the mortality rate for females in the age group of (70-79), known as the high-risk group, is calculated as \( \frac{17}{41} = 0.236 \) where 17 and 41 are the numbers of deaths and total patients, respectively. The recovery rate for the aforementioned group for females is also calculated as \( \frac{34+13}{72} = 0.653 \) where 34 and 13 are respectively the numbers of recovered and discharged patients following their personal satisfaction.

The cumulative proportion, required to perform the KS test, is directly calculated based on the mortality and recovery rates for both groups of females and males. They are obtained based on their previous cumulative proportion and the current one for each age group. For example, the cumulative proportion for females (70-79) years old is calculated as \( 0.308 + \frac{0.236}{1.560} = 0.460 \) where 0.308 is the cumulative portion for the age group of (60-69), 0.236 is the mortality rate for the current age group, and eventually, 1.560 represents the sum of mortality rates for all females’ age groups. The rest of the portions have been calculated following the above process and tabulated as mortality and recovery columns that are divided into patients’ gender.

The last column in each section is the difference between the cumulative proportion of females and males in each group. It is an absolute value of the difference between two portions for both gender groups. For example, for the aforesaid group, it is calculated \( |0.460 - 0.564| = 0.104 \). The maximum value of the above coefficient is known as KS stat and obtained as 0.127 for females and 0.246 for males demonstrated in the last row of Table 2.

The obtained values should be compared to the critical values of the KS test. Patients’ age for mortality is applicable in nine groups, so the critical value of KS \((0.95\%, 10)\)=0.409 is greater than the obtained value of 0.127. It shows there is no difference between the distribution functions of mortality for females and males. That is also categorized into 10 groups for recovery and the critical value of KS\((0.95\%, 10)\)=0.409 is greater than the obtained value of 0.246, so it shows there is no difference between females and males in terms of the distribution function of recovery rate.

4. Discussion

To demonstrate the results in a better way, the above calculations are also depicted in Figures 1 and 2, where the mortality and recovery rates for females and males are respectively shown by dashed and double lines. For each figure, the maximum absolute value of the difference between cumulative distribution functions is depicted next to an oval shape that is surrounding cumulative functions. They show the absolute value of the difference between females and males for the distribution function of mortality rate is 0.127 and for recovery rate is 0.246, both reveal that there is no significant difference between females and males. The final comparison is to investigate the dissimilarity between mortality and recovery rates for all patients. The last five columns of Table 2 demonstrate the same calculation processing results including the rates, cumulative portions, and eventually dissimilarities. They are also depicted in Figure 3. As shown in Table 2 and Figure 3 simultaneously, the KS stat is obtained as 0.430 which should be compared to KS\((0.95\%, 10)\)=0.409. Checking the obtained value and critical value reveals the mortality and recovery rates for all patients are different in terms of age, so it can be concluded that the mortality rate is different in age groups.
## Table 2. Stats calculated for females and males in mortality and recovery rates and their differences

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Mortality Rate</th>
<th>Cumulative Proportion</th>
<th>Recovery Rate</th>
<th>Cumulative Proportion</th>
<th>Rate</th>
<th>Cumulative Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Difference</td>
<td>Female</td>
</tr>
<tr>
<td>00-09</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>10-19</td>
<td>0.000</td>
<td>0.077</td>
<td>0.000</td>
<td>0.055</td>
<td>0.055</td>
<td>0.000</td>
</tr>
<tr>
<td>20-29</td>
<td>0.091</td>
<td>0.054</td>
<td>0.058</td>
<td>0.094</td>
<td>0.036</td>
<td>0.788</td>
</tr>
<tr>
<td>30-39</td>
<td>0.059</td>
<td>0.054</td>
<td>0.096</td>
<td>0.132</td>
<td>0.036</td>
<td>0.853</td>
</tr>
<tr>
<td>40-49</td>
<td>0.050</td>
<td>0.078</td>
<td>0.128</td>
<td>0.189</td>
<td>0.061</td>
<td>0.925</td>
</tr>
<tr>
<td>50-59</td>
<td>0.145</td>
<td>0.141</td>
<td>0.221</td>
<td>0.290</td>
<td>0.069</td>
<td>0.776</td>
</tr>
<tr>
<td>60-69</td>
<td>0.137</td>
<td>0.114</td>
<td>0.308</td>
<td>0.372</td>
<td>0.064</td>
<td>0.695</td>
</tr>
<tr>
<td>70-79</td>
<td>0.236</td>
<td>0.267</td>
<td>0.460</td>
<td>0.564</td>
<td>0.104</td>
<td>0.653</td>
</tr>
<tr>
<td>80-89</td>
<td>0.271</td>
<td>0.274</td>
<td>0.634</td>
<td>0.761</td>
<td>0.127</td>
<td>0.643</td>
</tr>
<tr>
<td>90-100</td>
<td>0.571</td>
<td>0.333</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
<td>0.429</td>
</tr>
<tr>
<td>Sum</td>
<td>1.560</td>
<td>1.393</td>
<td>Max</td>
<td>0.127</td>
<td>5.761</td>
<td>7.831</td>
</tr>
</tbody>
</table>

**Figure 1. Cumulative distribution functions of females and males in mortality rate**
5. Conclusion

Since perceiving the behavior of patients who infected to Covid-19 on mortality and recovery rates is very important to healthcare authorities, the Kolmogorov-Smirnov test has been utilized to investigate the similarity of the age-based distribution functions for females and males. The research has been conducted in the Iranian northern province of Guilan, where data for 1015 patients were available to conduct the study received from three hospitals that were designated to hospitalize the Covid-19 patients. The results of statistical analysis revealed that in terms of gender, the patients’ mortality, and recovery rates come from the same distribution functions if the age serves as a basis for categorizing patients, but they are different based on their age groups. In implication, it can be concluded that gender does not remain a significant contributing factor for mortality and recovery rates but age has a significant effect on the Covid-19 mortality and recovery rates. The results lead the healthcare authorities that they can manage Covid-19 patients regardless of their gender, but should be aware of their ages because the patient’s age play a significant role on the chance of mortality and recovery.

Researchers who are interested in working in this field are recommended to more focus on specific personal characteristics such as lifestyle, food, place of birth, and the other factors contributing to the mortality and recovery rates of the patients if collecting accurate data is possible.
6. List of Abbreviations

HIS  Health Information System.
KS  Kolmogorov-Smirnov.

7. Declarations

7.1. Data Availability Statement
The data presented in this study are available in article.

7.2. Funding
The author received no financial support for the research, authorship, and/or publication of this article.

7.3. Ethical Approval
No human or animal have been participated in the study process. The study was just conducted by analyzing the data fields gathered from HIS (Health Information System). In addition personal specifications such as patients’ phone numbers, address, etc. have not been received by author.

7.4. Acknowledgements
The author would like to express his great appreciation and gratitude for the support received from IT administrators of Poursina, Alzahra, and 17 Shahrvair hospitals and special thanks to Ms. Maedeh Pourmirza for providing the data in the required fields to conduct this research work.

7.5. Declaration of Competing Interest
The author declare that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

8. References


