**Final Research Paper – COVID-19 Symptoms and Demographics Severity Impact**

Erin Rooney

Colorado State University Global

MIS581 Capstone - Business Intelligence and Data Analytics

Dr. Mills

6/30/2024

**Abstract**

Since late 2019, COVID-19 has spread across more than 189 countries, resulting in over 774.88 million reported cases worldwide. The profound impact of severe COVID-19 cases on healthcare systems has been unmistakable, straining resources and personnel globally. This study aimed to empirically investigate how COVID-19 symptoms contribute to the severity of the disease, while focusing on assessing the potential for earlier medical intervention. Additionally, the research explored whether specific demographic factors, such as age and gender, correlate with an increased risk of severe COVID-19, aiming to inform targeted public health interventions for vulnerable populations. Analyzing data encompassing symptoms, age, and gender, the study employed advanced statistical techniques including k-means clustering and logistic regression models. These methods were selected to discern patterns that could indicate the severity of COVID-19. The effectiveness of the k-means clustering model was evaluated using metrics like the Silhouette Score and Adjusted Rand Index, providing insights into the separation of identified clusters. Meanwhile, logistic regression models were evaluated using accuracy and F1-Score metrics to understand their predictive performance in distinguishing severe cases from non-severe ones. Despite rigorous analysis, the study's findings yielded inconclusive results regarding the direct prediction of high-severity COVID-19 based on symptoms or demographic variables alone. However, notable indications suggested that age could play a significant role. This nuanced insight underscores the complexity of COVID-19 severity prediction, emphasizing the need for further investigation into how specific age groups may be predisposed to more severe outcomes. Moving forward, future research should aim to expand upon these preliminary findings by incorporating additional variables and employing more sophisticated modeling techniques. This approach could yield deeper insights into the complex interactions between symptoms, demographics, and COVID-19 severity, ultimately guiding more effective public health strategies and clinical interventions tailored to mitigate the impact of severe COVID-19 on vulnerable populations.

*Keywords:* COVID-19, Symptoms, Demographics, health interventions

**Final Research Paper – COVID-19 Symptom and Demographics Severity Impact**

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the virus responsible for the coronavirus disease 2019 (COVID-19), which spread to over 189 countries and has had over 774.88 million reported cases (Sheposh, 2024). Without knowing the severity of COVID-19, more individuals might seek medical attention than necessary causing hospitals to hit max capacity and medical professionals to get overbooked. Overbooking can lead to difficulty in obtaining a timely diagnosis or securing an appointment with a healthcare provider causing worsening symptoms and prolonging the risk of transmission. Phippen (2023) reported that in the United States, the average wait time to see a physician is 26 days, which further compounds the delay in receiving necessary medical care, especially when specialist appointments are involved. Conversely, knowing if a person’s symptoms indicate the need for early medical intervention could allow them to seek treatment promptly and prevent further complications minimizing their need for intensive care.

Certain demographics may be more likely to experience severe COVID-19 symptoms and require hospitalization or worse outcomes. By identifying which groups are at a higher risk of contracting severe COVID-19 symptoms, proactive and preventative approaches like wearing a mask or avoiding crowded public places can be taken.

One aim of my project is to develop a COVID-19 severity detection system based on the symptoms an individual might experience. By enabling individuals or healthcare professionals to input the symptoms and receive prompt feedback on whether patients should seek further medical attention, the system encourages a prescriptive approach to health monitoring and reduces unnecessary visits to healthcare facilities. This streamlined approach empowers individuals to take charge of their health while also alleviating the burden on healthcare providers, allowing them to prioritize critical cases. The simplicity and accessibility of this diagnostic tool has the potential to revolutionize healthcare delivery methods, offering alternative care options promptly while enhancing efficiency and resource allocation in the face of public health crises like COVID-19.

The second aim of this project is to conduct a comprehensive demographic analysis to identify populations at higher risk of experiencing severe COVID-19 outcomes. This knowledge is pivotal for devising targeted public health interventions and educational initiatives with the intent to prevent disease transmission. By understanding which demographics face heightened risks, healthcare and government resources can be strategically allocated to protect vulnerable groups and mitigate negative outcomes. Widespread broadcasting of this knowledge empowers the public to make informed decisions regarding social interactions and contact practices. Armed with awareness of their susceptibility to severe COVID-19, individuals can adopt preventive measures more effectively, such as adhering to recommended safety protocols and seeking timely medical assistance when needed. Moreover, targeted educational campaigns can emphasize tailored strategies for at-risk demographics, promoting behaviors that reduce transmission and protect vulnerable individuals.

**Objectives**

The objective of my project is to develop and deploy multiple machine-learning models aimed at predicting the severity of COVID-19 based on symptom profiles, while also examining how demographics influence disease severity. By leveraging advanced analytical techniques to analyze symptom data, these models aim to provide timely recommendations on the necessity of medical intervention. This approach ensures that individuals with severe COVID-19 symptoms receive prompt medical attention, optimizing healthcare resources, reducing associated costs, expediting patient care, and ultimately minimizing virus transmission. Effective allocation of healthcare resources through predictive modeling also plays a crucial role in mitigating risks associated with staff availability and capacity constraints within healthcare facilities. By understanding which demographics are more susceptible to severe COVID-19 outcomes, healthcare professionals and public health officials can devise targeted interventions. This proactive approach not only enhances patient outcomes but also empowers healthcare providers to prioritize care for high-risk groups. Furthermore, equipping the public with knowledge about their susceptibility to severe COVID-19 encourages proactive health behaviors and informed decision-making. By promoting awareness and tailored preventative measures, such as adherence to safety protocols and timely medical consultations, individuals can play an active role in safeguarding their health and reducing transmission risks.

**Overview of Study**

This research project and the models created will allow for better public health interventions and improved efficiencies in the healthcare system. This research contributes to the development of advanced diagnostic tools for COVID-19 severity detection, highlighting the importance of symptom and demographic data in managing public health responses. The study underscores the potential of machine learning in enhancing healthcare efficiency and resource allocation during public health crises.

**Research Questions and Hypotheses**

This project can be broken down into two research questions: the first focuses on identifying the most effective model for predicting COVID-19 severity based on an individual’s symptoms, and the second examines whether an individual’s demographics make them more susceptible to experiencing severe COVID-19 symptoms.

**Research Question 1**

What type of model between K-means cluster and multiple logistic regression is more effective at determining the severity of COVID-19 based on the symptoms the individual experienced? Understanding the severity of an illness, along with providing a prescriptive measure for whether patients should seek medical attention, can facilitate early identification, and treatment or recommend that patients remain at home, thereby reducing the risk of transmission to others.

The hypotheses for this research question are as follows:

**H1:** A clustering model is more effective than a regression model at determining the severity of COVID-19 based on the symptoms the patients experienced.

**H2:** A regression model is more effective than a clustering model at determining the severity of COVID-19 based on the symptoms the patients experienced.

**Research Question 2**

Are demographic factors, such as age and gender, significant indicators of if the patient will experience high-severity COVID-19? Understanding if demographic factors put an individual at higher risk of contracting a more severe form of COVID-19 is crucial. Identifying the groups at a higher risk allows for public health officials to intervene with preventative measures such as tailoring public health interventions. These interventions can be used to reduce incidences within these groups, leading to fewer severe cases and reduced mortality rates among high-risk populations. For this research question, I plan to use logistic regression analysis.

The hypotheses for this research question are as follows:

**H0:** Demographic factors, such as age and gender have a significant impact on COVID-19 severity.

**Ha:** Demographic factors, such as age and gender do not have a significant impact on COVID-19 severity.

**Literature Review**

Due to the timing and impact of COVID-19, there is no shortage of literature on the topic. Alizad et al. (2023) determined that gender was not a factor in determining the severity of COVID-19 symptoms, however, their study focused primarily on elderly patients. This research project also considers age, which could be a factor in the severity of one's COVID-19 severity. Iaccarino et al. (2020) found that more men die from COVID-19 or have a more severe virus than women, however, this study also included if the patient had any comorbidities such as obesity, or hypertension. With the dataset used for this research project, comorbidities are unknown, so when gender is tested, this might be overlooked. Abolfotough et al. (2022) found using logistic regression models that even though the count of positive cases showed more men, women were hospitalized at a larger rate proportionally. This would show that women are more likely to have severe symptoms resulting in hospitalizations. Finally, Abolfotough et al. (2022) found that men had a higher rate of in-hospital morbidity than women. This means that severity could be defined in multiple ways, and each definition could result in different answers. Defining if severity means hospitalization, a stay in the intensive care unit (ICU), or death could be crucial for further research. This transitions to the study Boudou et al. (2021) conducted where they differentiated hospitalization, ICU, and deaths as three levels of severity. In this study age was a more significant factor than gender for all three regression trees indicating that age plays a larger impact on determining a person’s COVID-19 severity than gender would, however, it is noted that male patients are three times the odds of requiring ICU admission and older men were more likely to have comorbidities than older women. This means that the results match that of Iaccarino et al.

In conclusion, literary research has revealed notable trends in the impact of COVID-19 across age and gender demographics. Specifically, women appear more likely to experience symptoms of COVID-19 than men, while men are more likely to be admitted into the ICU or die from severe symptoms. Although the data used in this study does not have that granularity, it can still inform if gender has a significant impact on the severity of COVID-19. Across all the studies that were reviewed, age emerged as an essential factor, with older patients consistently showing a higher likelihood of suffering severe COVID-19 symptoms. Although my research does not include data on comorbidities, which many studies in the field consider, it still provides valuable insights into the roles of gender and age in COVID-19 severity. The exclusion of comorbidities highlights a limitation of the dataset for this project. In future research it would be worth exploring primary sources for information on patients' COVID-19 severity that includes any pre-existing comorbidities like asthma, hypertension, or obesity. This would tie into the purpose of this research project to provide targeted public health recommendations by identifying high-risk groups, allowing health officials to develop more effective strategies for early intervention and resource allocation.

**Research Design**

**Methodology**

The dataset that has been selected for this research project is sourced from Kaggle, containing COVID-19 symptom data and demographic information of confirmed cases. This is considered a secondary dataset as it was compiled by another organization. It comprises of 10 binary features that indicate the presence of a symptom including fever, sore throat, body pain, and more. Additionally, There are 5 features to indicate the age ranges including ages 0-9 years, 10-19 years, 20-24 years, 25- 59 years, and 60 plus years of age. The dataset includes three binary features for gender including male, female, and transgender, as well as an indicator for if there was known contact with an individual who has contracted COVID-19. The data is quantitative in nature, presumably gathered through surveys or other self-reported methods. It is important to note that the dataset's documentation does not specify whether explicit consent was obtained for the collection and use of this data, highlighting a potential ethical consideration. No additional primary sources were included in this study.

**Methods**

Methods used for research question 1 will include developing and comparing multiple models and evaluating their effectiveness using metrics such as F1 score, Silhouette Score, and Adjusted Rand Index. Additionally, principal component analysis (PCA) will be used as a composite indicator for one of the multiple regression models and the k-means model. As suggested by Cartone & Postiglione (2021), a composite indicator can capture multidimensional concepts that are not adequately represented by a single indicator, potentially enhancing model efficiency. This approach will also help address moderate correlations discovered among some of the dataset's features shown in Figure 1, ensuring that the models are both robust and reliable. The final method I will employ is the normalization of the input variables. Normalization is crucial because it scales the data to a standard range, typically between 0 and 1, ensuring that all features contribute equally to the analysis. This is particularly important in clustering and regression models, as it prevents variables with larger scales from disproportionately influencing the results.

Testing these models will involve an iterative process of training and validation to determine which method provides the most accurate predictions of COVID-19 severity. By identifying the most effective model, we can improve clinical decision-making, prioritize healthcare resources more effectively, and provide timely interventions to those most in need. This research aims to contribute to the development of advanced diagnostic tools that can better manage the pandemic and its impacts on public health, so it is important that the models are not over-fit to the training data and rendered useless for the general public.

**Figure 1**

*Correlation Matrix for all Variables Within the Dataset.*

A graph with red and blue squares

Description automatically generated

*Note. Screenshot from the output in Jupyter*

To determine whether the null hypothesis is accepted for research question 2, the odds ratio will be used to assess if the likelihood of experiencing severe COVID-19 is significantly associated with specific demographic factors. The odds ratio will help evaluate the strength and direction of the association between demographic factors and the severity of COVID-19. If the odds ratio is significantly greater than 1, it would indicate that individuals within certain demographic groups are at a higher risk of severe COVID-19 outcomes. Conversely, an odds ratio of less than 1 would suggest a protective effect meaning it decreases the likelihood of severe COVID-19. This statistical analysis will provide critical insights that can inform targeted public health strategies and resource allocation, ensuring that interventions are directed where they are needed most. By doing so, healthcare systems can better protect vulnerable populations, improve patient outcomes, and optimize the use of medical resources during the pandemic.

**Limitations**

There are a few notable limitations within this dataset. The first is the lack of information regarding the comorbidities of individuals. When looking into the literature on COVID-19 severity, multiple articles included information on comorbidities and how they impact the severity of COVID-19. Without this additional information, some of the most at-risk individuals may be overlooked. Additionally, the absence of detailed demographics like race data or income level poses a limitation to this research. Race and income level could be significant factors in the severity of COVID-19 outcomes and would provide valuable insights into demographic impacts, including potential health disparities and the effectiveness of targeted interventions. The lack of this data restricts our ability to fully understand and address these disparities, potentially overlooking critical factors that contribute to the varying severity of COVID-19 among different groups. This limitation underscores the need for more comprehensive data collection in future studies to ensure a more inclusive and accurate analysis of COVID-19 severity across more diverse populations.

Another limitation of the dataset is the lack of a clear definition for "severe." This ambiguity can complicate the prescriptive nature of the research question regarding whether or not to seek medical attention. It is imperative to differentiate between varying degrees of severity such as requiring basic medical care, more drastic interventions like ventilator us, or outcomes such as death. By clearly defining these levels, targeted interventions could be developed to improve hospitalization and death rates. Clarifying what constitutes "severe" would enhance the dataset's utility, leading to more accurate predictions and better-informed public health strategies.

**Ethical Considerations**

Ethical considerations such as data privacy, data accuracy, and potential biases within the dataset are integral to the success of this research. To address privacy concerns, all personally identifiable information (PII) has been removed from the dataset, ensuring compliance with data protection regulations. Data accuracy is crucial for the integrity of machine learning models. Errors in the input variables can lead to incorrect model outcomes, resulting in recommendations for inappropriate actions that may worsen symptoms and potentially threaten lives. Such inaccuracies in the data can compromise the effectiveness of interventions, misguide healthcare professionals, and harm public health efforts. The data has been scrubbed for any extreme outliers that may indicate inaccuracies in the reported symptoms. Given that most input variables are binary indicators, the impact of a few inaccuracies is expected to be minimal and therefore will not be considered a moderate or severe risk. Addressing potential biases, particularly in demographic representation, is critical to prevent skewed outcomes. The dataset has been checked to ensure a diverse representation of genders and ages, reflecting the broader population. This diversity is essential for generalizing the findings of the research.

The initial exploration into COVID-19 severity underscores the critical need for effective predictive models and demographic analysis. Understanding which symptoms and demographic factors contribute most significantly to severe outcomes will inform targeted public health interventions, optimize healthcare resource allocation, and empower individuals to manage their health proactively amidst global health crises like COVID-19. These insights pave the way for developing advanced diagnostic tools and strategies that enhance healthcare efficiency and mitigate the impact of the pandemic on vulnerable populations.

**Findings**

**Finding for Research Question 1**

What type of model between K-means cluster and multiple logistic regression is more effective at determining the severity of COVID-19 based on the symptoms the individual experienced? To answer this question, 3 total models were evaluated: one k-means cluster and two regression models. The k-means model and the first regression model both used Principal Component Analysis (PCA) for dimensionality reduction and to combat the moderate correlation among some predictor variables. To determine the number of principal components to use, a plot for the cumulative explained variance by the number of components was created. As illustrated in Figure 2, the explained variance reaches 100% at about 14 principal components, indicating that adding more principal components would not improve the explained variance after 13. Therefore, it was determined best to iterate the creation of the k-means models with 11, 12, and 13 principal components to see which performed the best. Results for the three models showed that using 12 principal components was sufficient and using 13 principal components did not improve the outcomes, so the final models used 12 principal components.

**Figure 2**

*Cumulative Explained Variance for Principal Components*

A graph with a blue line

Description automatically generated

*Note. Screenshot from the output in Jupyter*

Next, the Elbow Method was used to determine the optimal number of clusters (k) for the K-means model. The elbow graph displays the within-cluster sum of squares (WCSS) values on the y-axis corresponding to various values of K on the x-axis. The optimal K value is identified at the point where the graph forms an elbow. The graph shown in Figure 3 suggests an optimal k value of 3 or 4. The choice was made to use 4 clusters given that the target variable for this dataset has 4 ordinal levels.

**Figure 3**

*Elbow Method to track inertia by the number of clusters*

A graph with a line

Description automatically generated

*Note. Screenshot from the output in Jupyter*

The Silhouette score and Adjusted Rand Index were used to determine the effectiveness of the k-means model. The results for both metrics can be found in Figure 4. The silhouette score of 0.25 indicates fair clustering with some separation among the different clusters, but not enough to result in high predictive power. According to Chacón & Rastrojo (2020), the Adjusted Rand Index measures the degree of agreement between the clusters, accounting for random distribution. The resulting Adjusted Rand Index of -0.00 indicates that the points are essentially random. Overall, while the cluster model works for separation, it is not effective at predicting if symptoms can accurately identify the severity of an individual's COVID-19.

**Figure 4**

*Silhouette Score and Rand Index for the k-means model.*

A screenshot of a computer

Description automatically generated

*Note. Screenshot from the output in Jupyter*

In addition to the k-means model, two multiple logistic regression models were built. The first model used PCA for dimensionality reduction, following the same steps as the clustering model (see Figures 2 and 3 for details). After fitting the PCA to the training data, logistic regression was performed and assessed with the testing data. The results of the logistic regression using PCA are presented in Figure 5.

**Figure 5**

*Results of the Multiple Logistic Regression Model Using Principal Component Analysis.*

A screenshot of a computer screen

Description automatically generated

*Note. Screenshot from the output in Jupyter*

These results indicate an accuracy of 75%, however, the model failed to identify any cases of severe COVID. This deficiency is underscored by the F1-score which indicates that the model did not correctly classify any instances in the test dataset as severe (1). Ethically, false negatives carry significantly more weight because misjudging high severity for lower severity levels could potentially lead to fatal consequences if necessary medical interventions are delayed. Ensuring the model's ability to accurately identify severe cases of COVID-19 is crucial for guiding appropriate medical responses and interventions.

The second regression model was intentionally executed without PCA. As depicted in the results in Figure 6, the overall accuracy of the model decreased to 47%. Though the accuracy decreased from the regression model with PCA, notable improvement was observed in predicting severe cases of COVID-19, as indicated by the F1-score for severe cases (1). This metric highlights the model’s effectiveness in correctly identifying instances where medical intervention could be an urgent matter. Ethically, the ability to minimize false negatives (cases where the model fails to identify severe COVID-19) holds paramount importance. By focusing on optimizing the prediction of severe cases without relying on dimensionality reduction through PCA, this approach underscores the model’s capacity to prioritize clinical urgency over overall model accuracy metrics alone. Due to this, it was determined that the final regression model with 47% accuracy outperformed the initial regression model.

**Figure 6**

*Results of the Multiple Logistic Regression Model Without PCA*

A screenshot of a computer

Description automatically generated

*Note. Screenshot from the output in Jupyter*

Based on the results presented in research question 1, none of the models effectively determines the severity of COVID-19 and none would be appropriate for deployment or potential use within the healthcare system. For this, we neither accept nor reject the null hypothesis as no models exceeded the others. It may be beneficial to include additional models, such as decision trees or artificial neural networks, as they might be more effective in predicting severe cases of COVID-19.

**Finding for Research Question 2**

Are demographic factors, such as age and gender, significant indicators of if the patient will experience high-severity COVID-19?

For investigation into this research question, a logistic regression model was utilized to assess the impact of age and gender demographics on high-severity COVID-19 cases. Detailed results of the initial model can be found in Figure 7. Initially, the significance of predictor variables was evaluated using their respective p-values. Upon review of Figure 7, it was observed that none of the predictor variables achieved statistical significance, although age exhibited relatively lower p-values. In light of these findings, a subsequent model was developed, excluding gender and focusing solely on age as a predictor. The results of this refined demographics model are detailed in Figures 8 and 9. These figures provide comprehensive insights into the relationship between age and high-severity COVID-19 outcomes, shedding light on whether age alone can serve as a predictive factor.

**Figure 7**

*Results of the First Logistic Regression Model on Demographics*

A screenshot of a computer

Description automatically generated

*Note. Screenshot from the output in Jupyter*

**Figure 8**

*Results of the Second Logistic Regression Model on Demographics Excluding Gender*

A screenshot of a computer

Description automatically generated

*Note. Screenshot from the output in Jupyter*

**Figure 9**

*Results of the Odds Ratios for the Second Logistic Regression Model on Demographics Excluding Gender*

A number on a white background

Description automatically generated

*Note. Screenshot from the output in Jupyter*

In Figure 8, only Ages 25-59 demonstrate a weak statistical significance, with a p-value of 0.075. Although this value does not meet the conventional significance threshold of 0.05 to reject the null hypothesis, it suggests a potential weak association between age and COVID-19 severity. This finding indicates that individuals aged 25-59 may merit further analysis in future studies exploring age-related factors in COVID-19 severity.

Figure 9 illustrates that the odds ratios for all variables were less than 1, indicating that none of the input variables are statistically significant in predicting whether an individual will experience a severe case of COVID-19. This emphasizes our inability to reject the null hypotheses related to age and gender as predictors of severity.

Overall, concerning our research question, we cannot assert that age or gender has a statistically significant impact on COVID-19 severity, thereby leading us to fail to reject the null hypothesis. However, the marginally significant p-value for Age 25-59 (0.075) suggests that additional investigation may be warranted to delve deeper into potential associations between specific age groups and severity outcomes. These findings underscore the complexity of COVID-19 severity prediction and highlight the need for a comprehensive exploration of contributing factors beyond age and gender.

**Conclusion**

Based on the findings in this study, none of the models effectively predict the severity of COVID-19, preventing their deployment within healthcare systems. Despite achieving up to 75% accuracy the logistic regression models failed to identify any severe COVID-19 cases, as indicated by an F1-score of zero for severe cases. This limitation raises the dangerous ethical concern of the risk of false negatives, where severe cases are misclassified, potentially delaying necessary medical interventions, and leading to more adverse outcomes.

The exploration of demographic factors such as age and gender did not yield statistically significant associations with COVID-19 severity, suggesting that these variables alone may not reliably predict high-severity outcomes. However, including other demographics, such as race, could enhance predictive accuracy.

Subsequently, further research is recommended on comorbidities to improve model accuracy and incorporate additional relevant factors. This approach could better inform public health strategies and clinical decision-making during pandemics, ensuring timely interventions and optimizing resource allocation to reduce severe outcomes. Continued investigation into more comprehensive datasets and advanced modeling techniques will be crucial in addressing these challenges.

**Recommendations**

Further exploration could benefit from more granularity in symptom reporting, particularly by distinguishing between different severity levels for each symptom. For example, categorizing symptoms like cough or fatigue on an ordinal scale (e.g., mild, moderate, severe) or using a numerical rating (e.g., 1-5) could provide more nuanced insights into their impact on COVID-19 severity. This level of granularity would allow for differentiation between mild symptoms that may not hinder day-to-day activities and severe symptoms necessitating immediate medical attention. Likewise, while age buckets are useful for anonymizing data, employing age as a continuous variable could offer a clearer understanding of any linear correlation between age and COVID-19 severity. A continuous age variable would provide a better input for the model to improve the efficiency and deliver a more precise analysis of age-related trends and severe outcomes.

Additionally, accessing a more comprehensive dataset from respected sources like the Centers for Disease Control and Prevention (CDC), the World Health Organization (WHO), or other government health organizations could enhance the accuracy and predictive power of future studies. Such datasets typically include a broader range of variables, such as comorbidities and detailed symptom profiles, which are crucial for capturing the complexity of COVID-19 severity factors. Incorporating these improvements would not only refine predictive models but also strengthen the validity and applicability of findings, thereby supporting more informed public health strategies and clinical decision-making in combating COVID-19

Finally, addressing the ambiguity of what defines severity could largely impact the effectiveness of the model(s). Differentiating if severity means hospitalization, a stay in the intensive care unit (ICU), or death is crucial, as it shapes public health messaging and interventions, ensuring that resources are targeted appropriately to those at greatest risk. Understanding these distinctions can enhance the precision of health guidelines and improve patient outcomes across the globe.

**References**

Abolfotouh, M. A., Musattat, A., Alanazi, M., Alghnam, S., & Bosaeed, M. (2022). Clinical characteristics and outcome of Covid-19 illness and predictors of in-hospital mortality in Saudi Arabia. BMC Infectious Diseases, 22(1). https://doi.org/10.1186/s12879-022-07945-8

Alizad, G., Ayatollahi, A. A., Shariati Samani, A., Samadizadeh, S., Aghcheli, B., Rajabi, A., Nakstad, B., & Tahamtan, A. (2023). Hematological and Biochemical Laboratory Parameters in COVID-19 Patients: A Retrospective Modeling Study of Severity and Mortality Predictors. BioMed Research International, 2023, 7753631. https://doi.org/10.1155/2023/7753631

Boudou, M., ÓhAiseadha, C., Garvey, P., O’Dwyer, J., & Hynds, P. (2021). Modelling COVID-19 severity in the Republic of Ireland using patient co-morbidities, socioeconomic profile and geographic location, February to November 2020. Scientific Reports, 11(1), 1–11. <https://doi.org/10.1038/s41598-021-98008-6>

Chacón, J. E., & Rastrojo, A. I. (2020). Minimum adjusted Rand index for two clustering’s of a given size.

Chen, Y., Wu, J., Ma,J., Zhu H., Li, W., & Gan, Y. (2022). The mediating effect of media usage on the relationship between anxiety/fear and physician-patient trust during the COVID-19 pandemic. *Psychology & Health, 37*(7), 847–866. https://doi.org/10.1080/08870446.2021.1900573

Iaccarino, G., Grassi, G., Borghi, C., Carugo, S., Fallo, F., Ferri, C., Giannattasio, C., Grassi, D., Letizia, C., Mancusi, C., Minuz, P., Perlini, S., Pucci, G., Rizzoni, D., Salvetti, M., Sarzani, R., Sechi, L., Veglio, F., Volpe, M., & Muiesan, M. L. (2020). Gender differences in predictors of intensive care units admission among COVID-19 patients: The results of the SARS-RAS study of the Italian Society of Hypertension. PloS One, 15(10), e0237297. <https://doi.org/10.1371/journal.pone.0237297>

Phippen, E. (2023, January 5). New Survey: Physician appointment wait times are getting longer. *Washington State Hospital Association.*

Sheposh, R. (2024). Coronavirus Disease 2019 (COVID-19). Salem Press Encyclopedia of Health.