Final Report

Deeya Shukla

Introduction

Covid-19 vaccines offer protection against the coronavirus, an infectious disease. It works by increasing the body's natural immunity against the virus (Department of Health, 2021). There are three approved vaccinations in Australia: Pfizer, AstraZeneca, and in August 2021, the Moderna vaccine was approved for use by the Therapeutic Goods Administration (Therapeutic Goods Administration, 2021). Although vaccines are safe to use, they can cause adverse reactions. This report will focus on the adverse reactions caused by the Pfizer vaccine compared to the Moderna vaccine. With the new Moderna vaccine being introduced in Australia, hospitals, doctors and nurses should be aware of the most commonly caused reactions and know the chance of getting a reaction and being hospitalised. This information will assist them in preparing treatment for the most common reactions to the Moderna vaccine. Moreover, by knowing the chances of a reaction after the vaccine, the results will assist the general public in deciding which vaccine is better suited for them.

Dataset

The dataset is sourced from Kaggle; however, it was scrapped from the Vaccine Adverse Event Reporting System (US) (VAERS) (Taukir Chowdhury, 2021). VAERS is utilised by the Food and Drugs Administration (FDA) and the Center for Disease Control and Prevention (CDC); therefore, it is a credible source (VAERS, 2021). Furthermore, the dataset was updated on the 11th of August 2021, and it contains information from January to August 2021. Moreover, it has more than 600,000 rows and 52 columns; hence it is sufficiently large to draw accurate and relevant conclusions.

I was motivated to choose this dataset because the topic of covid-19 vaccines is relevant in the current world environment. Moreover, with the recent reveal that the Moderna vaccine was safe for use in Australia, I was curious to find out more about this vaccine. I believed that the information from the results would allow me to make an informative decision on which covid-19 vaccine I should get. When looking into the data description, I was intrigued by the different attributes in the dataset, such as the symptoms, hospitalisation rate, days spent in the hospital, mortality rate and medical history. At first, I wanted to explore the correlation between symptoms and medical history; however, there was a significant amount of missing information. This may have occurred because not every patient has a medical history, or some may choose to leave it out. If I continued with this plan, it might imply an incorrect correlation between medical history and the type of reactions. Therefore, I changed the focus of my presentation towards the question, "What are the adverse reactions caused by the Pfizer vaccine in comparison to the Moderna vaccine?". Furthermore, the sub-question was "How do the vaccines impact stakeholders?" stakeholders included patients, hospitals, nurses, doctors and the general public that wanted to get vaccinated.

In conjunction with the dataset, Our World Data, an online source, was utilised to create one graph (Our World Data, 2021). Our World Data provided information about the total number of dose-one vaccinations administered for each vaccine. In addition, this source, combined with the dataset, assisted in creating a pie chart for the chance of reaction because the dataset was collected from the population size provided by the source.

Method/experimental set-up

To create most of the visualisation, Python3 was utilised in JupiterNotebook with a variety of libraries. I chose to use Python over other programming languages and software because the various libraries enabled me to create a range of graphs. Furthermore, unlike excel, Python was able to open my CSV file. Additionally, by utilising Python, I will learn a programming language that I haven't learned or used before, which is both exciting and challenging.

I imported the CSV file containing the dataset into a new data frame to begin the visualisations, using the Pandas library. Next, the data had to be cleaned, and I did this by removing rows with duplicate information. Moreover, I extracted the most relevant columns into the main data frame, and I did this in order to focus on the relevant attributes. Later in the coding process, as the focus changed, I added extra columns to the primary data frame, and the data frame was re-cleaned for any rows which contained null or unimportant values. To ensure I had controlled variables, I filtered the data to only contain information for the first dosage. Hence all the data is related to the first dosage of both vaccines.

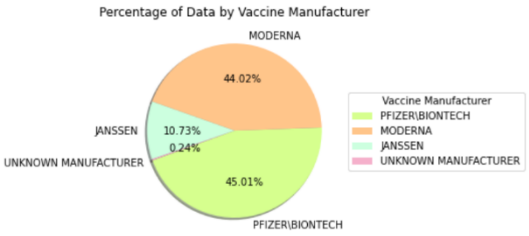


Figure 1: Split of data for each vaccine

The data contained information about Moderna, Pfizer, Johnson and Johnson and unknown vaccination manufacturer information. To understand the split between the data, I created a pie chart using the matplotlib library, refer to figure 1. The pie chart showed that the data was primarily focused on the Moderna and Pfizer vaccine. This worked out in favour of the question as these two vaccines are the most important to the relevant stakeholders in Australia.

For the first visualisation, I wanted to answer the main question directly: what are the adverse reactions caused by the vaccine? In order to determine the best graphing method, I used the Numpy library to count the values for each symptom. The results revealed more than 5000 symptoms; therefore, I chose only to show the top 10 most common reactions within both vaccine groups by creating a bar graph. At first, I decided to combine the information onto one chart, but when I quickly realised that the top 10 symptoms were different for both groups, I created two separate bar graphs. Once created, it became apparent that the x-axis of the bar graphs showed that the chance of getting a reaction from the Moderna vaccine was slightly higher than the Pfizer vaccine. To present this result, I decided to create two simple bubble charts that would be easily interpretable by the proportion of each bubble. I did this in an infographic website called, Venngage, which had the option to create bubble charts by inputting the total number of patients with each symptom. However, when looking at the new visualisations, the results I had hoped to show through the bubble proportions actually became more inconspicuous than before, refer to figure 2. The proportion of the bubbles between the two vaccine groups was indifferent. Therefore, I decided to utilise the initial bar graph.

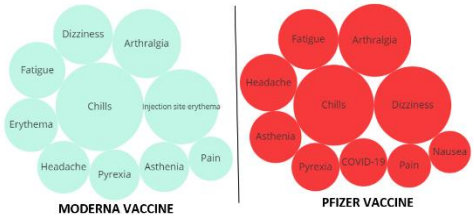


Figure 2: trial bubble charts

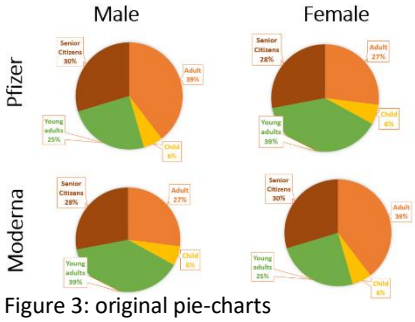


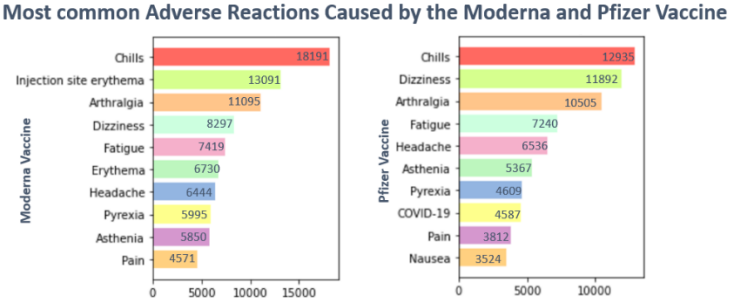
Figure 3: original pie-charts

For the second visualisation, I wanted to present the probability of getting a reaction for each age group. To do this, I split the age groups by the related demographic, e.g. elderly citizen, young adults etc. However, as this wasn't split evenly, I had to change the age ranges to be of equal sizes. When this visualisation was finally finished, my tutor pointed out that this was not the chance of getting a reaction, but rather it showed which age group had the highest/lowest chance of getting a reaction, refer to figure 3. Which may have been impacted by the bias from the total number of results reported for each age range. To improve this and get back on track to finding the probability of getting a reaction, I was able to find an external source that presented the number of people that got the first dose of each vaccine from January to August. Utilising this data in conjunction with the number of people who faced an adverse reaction to the respective vaccines. I was unsure of choosing between a stacked column graph or a pie chart to show the results. When I created a column graph, the proportion of those that got a reaction could not be seen. Therefore, I opted to make a pie chart and explode the segment, which showed the percentage of reactions vs no-reactions to visualise the difference clearly.

For the last two visualisations, I wanted to present the hospitalisation rate and time spent in the hospital after an adverse reaction. To show the hospitalisation rate, I opted to create a pie-chart because a column graph could not clearly visualise the difference between the two vaccines. Afterwards, I removed the rows with any null values in the hospitalisation column. Next, I used the seaborn library to plot the data as a density distribution graph because it shows the days spent at the hospital over a continuous x-axis. I then created two new data frames to filter the data by age and the time spent in the hospital for each vaccine group.

Project 1 Final Presentation Results and Analysis

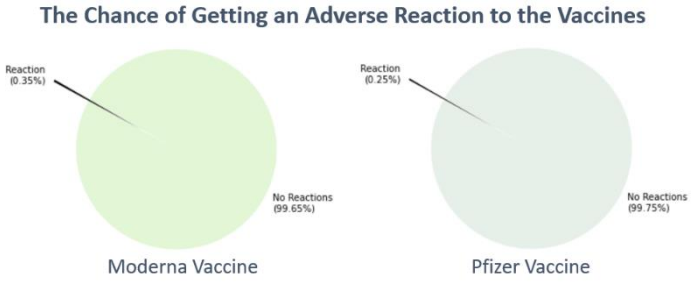
Visualisation 1:



This bar graph shows the most common reactions caused by the Moderna and Pfizer vaccine. It can be seen that most of the reactions are common among the two vaccines, such as chills, headache, arthralgia (joint pain), fatigue and dizziness. However, for the Moderna vaccine, the second most common symptom is Injection site erythema (swelling and pain around the injection site). Whereas one of the top-10 symptoms for the Pfizer vaccine is COVID-19; this is interesting because that is why the vaccine is taken. This may be a result of the population getting vaccinated whilst having community cases of covid-19. If this was to apply to a different demographic such as Adelaide, that symptom might not be prevalent because there are no community cases, therefore, it is improbable it would occur. Additionally,

another point of difference is the volume of each reaction, measured by the x-axis on both graphs. It can be seen that for the Moderna vaccine, there is a slightly higher chance of getting the reactions in comparison to the Pfizer vaccine. This comparison is further explored in the second visualisation.

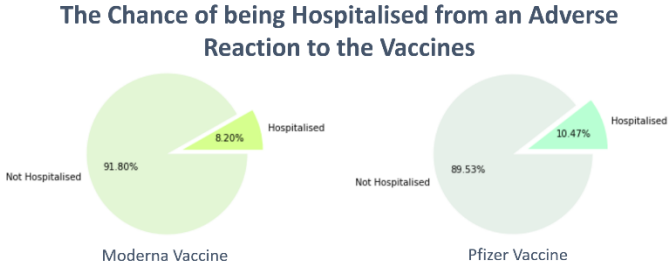
Visualisation 2:



Visualisation 2 is the chance of getting an adverse reaction to the vaccines. The pie charts suggest that the Moderna vaccine has a 99.65% chance of causing no reaction and a 0.35% chance of causing a reaction. This is slightly higher than the Pfizer vaccine, which has a 0.25% chance of causing an adverse reaction. The chances of getting a reaction to either vaccine is significantly low as suggested by the data. If there was more time this hypothesis could be tested by a significance test.

This data was created utilising two different sources; therefore, the validity of the results had to be compared to previous findings. A report produced by the CDC in November 2020, at the start of vaccinations in the US, stated that "(0.2%) adverse events after receipt of Pfizer COVID-19 vaccine " (Shimabukuro, 2020). The data found in this pie chart shows a similar value of 0.25% for the dataset ranging from January to August 2021. Therefore, although the two sources were utilised, it can be gathered that the results are moderately accurate.

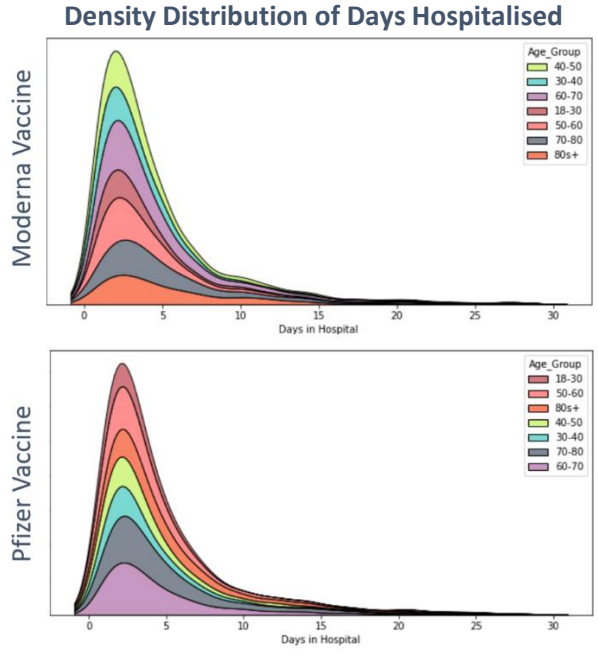
Visualisation 3:



This graph shows the chance of being hospitalised from an adverse reaction to the vaccines. For the Moderna vaccine, the chance of being hospitalised after an adverse reaction is 8.2%, whereas the probability is slightly higher at 10.47% for the Pfizer vaccine.

These results can assist hospitals in recognising how much of the vaccinated population will require treatment due to an adverse reaction. For example, suppose 1 million people in Adelaide get vaccinated with the Pfizer vaccine. In that case, roughly 2500 will get a reaction, according to the results, and from this, approximately 250 people will get hospitalised over that vaccination period. This amount of patients should be easily accommodated for considering there are more than 10 hospitals in the Adelaide region.

Visualisation 4:



The final visitation is a density distribution graph for the Days Hospitalised for each vaccine. For both graphs, it can be seen that the mean time spent in hospital is three days. The graphs show that those aged from 40-50 that get an adverse reaction to the Moderna vaccine have the highest chance of being hospitalised for any period of time. Whereas, for the Pfizer vaccine, those aged between 18-30 years old have the highest probability of being hospitalised for any period of time.

These results don't correlate to the prediction: as the age range increases, the likelihood of being hospitalised increases. These graphs instead have no pattern to which age groups are likely to be hospitalised most. I predict this may be because those age ranges have more results than the older age ranges. Additionally, it might be because elderly patients didn't know how to report their adverse reaction.

Conclusion

The vaccines cause similar reactions such as chills, headache, arthralgia (joint pain), fatigue and dizziness. However, there is a slightly higher chance of getting a reaction from the Moderna vaccine. Therefore, given the evidence from this data, the general public should be advised that the chance of an adverse reaction from the Pfizer vaccine is 0.25% which is comparatively lower than that of the Moderna vaccine (0.35%). Therefore, these probabilities aren't significant enough to stop the general public from getting vaccinated. Instead, the chances are so low that COVID-19 vaccinations are encouraged.

Discussion

In future, when there is more time allowed, the significance for the chance of getting an adverse reaction to either vaccine should be tested by a significance test to prove if those values are really inconsequential. Moreover, an in-depth visualisation could be created for the second visualisation if this dataset, combined with the dataset used in the external source, is split by age to predict which age range is most vulnerable to getting an adverse reaction. Finally, it would also be interesting to explore the correlation between the rate of hospitalisation and medical history using a more extensive dataset that doesn't contain missing information.

References

Department of Health, 2021. *How do COVID-19 vaccines work?.* [Online]
Available at: https://www.health.gov.au/initiatives-and-programs/covid-19-vaccines/learn-about-covid-19-vaccines/how-do-covid-19-vaccines-work
[Accessed 9 September 2021].

Our World Data, 2021. *Coronavirus (COVID-19) Vaccinations.* [Online]
Available at: https://ourworldindata.org/covid-vaccinations?country=USA
[Accessed 10 September 2021].

Shimabukuro, T., 2020. *Allergic Reactions Including Anaphylaxis After Receipt of the First Dose of Pfizer-BioNTech COVID-19 Vaccine — United States.* [Online]
Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7808711/
[Accessed 10 September 2021].

Taukir Chowdhury, 2021. *COVID-19 Vaccine Adverse Reactions (VAERS) Dataset.* [Online]
Available at: https://www.kaggle.com/landfallmotto/covid19-vaccine-adverse-reactions-vaers-dataset
[Accessed 9 September 2021].

Therapeutic Goods Administration, 2021. *TGA provisionally approves Moderna's COVID-19 vaccine.* [Online]
Available at: https://www.tga.gov.au/media-release/tga-provisionally-approves-modernas-covid-19-vaccine
[Accessed 9 September 2021].

VAERS, 2021. *VAERS DATA USE GUIDE.* [Online]
Available at: https://vaers.hhs.gov/docs/VAERSDataUseGuide_November2020.pdf
[Accessed 9 September 2021].