**VAERS Data**

Lincoln Brown

DSC530-T301

Dr. Parajulee

Bellevue University

**Statistical Hypothesis/Question:**

I became aware of the Vaccine Adverse Event Reporting System (VAERS) because of the COVID-19 pandemic and the vaccines developed to fight it. The worldwide focus on this event and the reaction from governments has magnified public attention to the adverse events caused by the COVID-19 reactions.

The data I used in the analysis includes cases up to July 2nd, 2021. The number of adverse events reported in 2020 was 48,901 and as of July 2nd, 2021 this number rose to 406,001, roughly an 830% increase. My intention in examining this data was to look for an explanation as to why the number of adverse events reported increased so dramatically. Did more elderly people receive a COVID-19 vaccination and have an adverse event that was unrelated to the vaccine, but needed to be recorded because of the timeliness of it? Were more vaccines administered due to COVID-19 than were administered in previous years?

Many of these original questions are unable to be answered with the VAERS data alone, so I decided to focus on the question related to elderly people receiving a vaccine, when they previously were not likely to have been vaccinated.

I used this data to form the hypothesis that the average age between adverse events in 2020 was different from the average age in 2021.

**Outcome of your EDA**

The outcome of my EDA resulted in a rejection of the null hypothesis that the difference in age between 2020 and 2021 was due to chance. With 99% confidence, we can assume that the difference in average age between 2020 and 2021 was not due to chance. The distribution of ages in 2021 suggests that adverse events were experienced by people of ages between 20 and 115. To me, this suggests that there is a possibility that these vaccines should receive more attention for safety from the FDA, but I want to stress that the VAERS data is not conclusive.

During my analysis, I also discovered that females are much more likely to experience an adverse event. In both 2020 and 2021, there were more than double the adverse events concerning females than there were concerning males.

**What do you feel was missing during the analysis?**

I really wish that I would have looked for more numerical information to include in my analysis. Age was the only numeric variable that I found useful to work with and I think that if I had more numerical data, I would have been able to create better regression models.

**Were there any variables you felt could have helped the analysis?**

I would have liked to include more information about the symptoms and places where the vaccine was received, but my experience limited my ability to work with these data effectively. I also would have liked to compare the number of vaccinations by state with the number of adverse events reported.

**Were there any assumptions made you felt were incorrect?**

I assumed normal distribution for the ages in 2020, and looking at the data, I do not believe this was a correct assumption.

**What challenges did you face, what did you not fully understand?**

I did not fully understand the challenges that I faced working with categorical data. Furthermore, the VAERS data is a voluntary reporting system. It does not provide complete or thorough accounting for all adverse events and there is the potential for duplicate information. The data had a lot of missing values, which made it tricky to work with. All of my analysis was conducted with these limitations in mind and I do not think that this analysis should be applied to any medical decision making. I believe that the pandemic increased awareness of VAERS and likely resulted in an increase in reports that normally would not have been made for other vaccines.

**Other thoughts**

I really enjoyed this project, it was definitely a challenge and I feel like I gained valuable experience working with large datasets. Although the results of the analysis should be taken with caution, I think that an 830% increase in reported adverse events is worthy of further investigation with a more reliable dataset.