Final Project VAERS Database

Lincoln Brown

2021-07-16

Dataset Information

Introduction

The COVID-19 pandemic has effected the entire globe and pushed the development of new vaccines designed to lessen the severity of COVID-19 symptoms. The vaccines were developed under project Warp Speed, which was designed to get the vaccines to the public as fast as possible. The vaccines underwent clinical trials rigorous enough to meet the Food and Drug Administration's requirements for Emergency Use Authorization (EUA).

Given the vaccines rapid roll out and their current pending full approval from the FDA, it is natural for people to be hesitant to take a newly developed vaccine, even with the gravity of COVID-19. The government has been monitoring adverse reactions with the Vaccine Adverse Event Report System (VAERS) since 1990 when it replaced Monitoring System for Adverse Events Following Immunization (MSAEFI).

VAERS is co-managed by the Center for Disease Control and Prevention (CDC) and U.S. Food and Drug Administration (FDA). VAERS is not designed to detect if a vaccine caused an adverse event, but it can identify unusual or unexpected patterns of reporting that might indicate possible safety problems requiring a closer look (VAERS, n.d.)

VAERS accepts reports from people who have received vaccines and experienced adverse effects or from healthcare providers who are required by law to report:

- Any adverse event listed in the VAERS Table of Reportable Events Following Vaccination that occurs within the specified time period after vaccinations
- An adverse event listed by the vaccine manufacturer as a contraindication to further doses of the vaccine

Knowingly filing a false VAERS report is a violation of Federal law (18 U.S. Code § 1001) punishable by fine and imprisonment (VAERS, n.d.).

There is a chance for duplicate reports if someone reports an adverse reaction on their behalf or on behalf of a family member and the healthcare provider also reports it (VAERS, n.d.).

The number of adverse reactions reported to VAERS for the COVID-19 vaccines is 10,875 in 2020 and 35,4451 as of July 9th, 2021. Of these events reported, 16 were associated with deaths in 2020 and 5718 were associated with deaths in 2021.

Some rare adverse side effects of the vaccines have been reported, such as Guillain-Barre Syndrome, Thrombosis with thrombocytopenia syndrome (TTS), myocarditis, and pericarditis.

Adverse events reported to VAERS fall within a specific time limit of receiving the vaccine and do not necessarily mean that they were caused by the vaccination, this is especially true of deaths.

Research questions

This paper will be exploring 5 questions regarding the VAERS database:

1. Of the deaths that occurred, what was the average age?
2. How many of the deaths involved a vaccine given in a nursing home?
3. Does the type of vaccine have any correlation with the death rate.
4. What effect did age have on adverse events?
5. What role did sex play in adverse events?
6. What timeline do the vaccine adverse reactions follow? Is there a bloom of adverse reactions around the time they were released?
7. Is there evidence of the more vaccinated states reporting more adverse reactions?
8. Were more adverse reactions reported for the first or second dose of the COVID-19 vaccines.

Approach

My hypothesis is that most of the deaths are in the elderly and are associated with vaccines given in nursing homes. This approach will look at several of the variables related to life threatening events, where the vaccine was administered, and the type of vaccine given. Looking at variables such as age and sex will help give a better understanding of those most frequently affected by adverse reactions. I also want to explore the number of COVID vaccines given in relation to other vaccines to see if the number of vaccines given is more than previous years.

This project is not intended to and will not be able to provide any sort of proof that the vaccines are harmful; because reported adverse reactions are required to be submitted after an event, whether the vaccine caused it or not. These reports are indications that these reactions occurred within a specific time frame of receiving a vaccine, but do not imply causation. I will create a regression model to see how age, sex, and where the vaccine was administered impacts the likelihood of death. I also will check for confounding variables such as the type of vaccine and previous allergic reactions.

How your approach addresses (fully or partially) the problem. The problem will only be able to partially be addressed because the data is not conclusive proof one way or another. There are also ethical problems with using these data for concrete evidence in either direction. My intention is to observe the adverse reactions in a scientific light to see if there are explanations for the adverse reactions other than the COVID-19 vaccines.

Data (Minimum of 3 Datasets - but no requirement on number of fields or rows)

VAERS data are distributed in three datasets, VAERSVAX, VAERSDATA, and VAERSSYMPTOMS. I will be using the datasets for the years 2019, 2020, and 2021.

In addition, I will be using 3 datasets for the COVID virus cases and deaths provided by USA Facts.

- COVID Confirmed Cases
- COVID County Population
- COVID Deaths

```
data19 <- read.csv("2019VAERSDATA.csv")
symp19 <- read.csv("2019VAERSSYMPTOMS.csv")
vax19 <- read.csv("2019VAERSVAX.csv")
data20 <- read.csv("2020VAERSDATA.csv")
symp20 <- read.csv("2020VAERSSYMPTOMS.csv")
vax20 <- read.csv("2020VAERSVAX.csv")
data21 <- read.csv("2021VAERSDATA.csv")
symp21 <- read.csv("2021VAERSSYMPTOMS.csv")
```

```r
vax21 <- read.csv("2021VAERSVAX.csv")

confirmed <- read.csv("covid_confirmed_usafacts.csv")
population <- read.csv("covid_county_population_usafacts.csv")
deaths <- read.csv("covid_deaths_usafacts.csv")


print(dim(data19))
```

```
## [1] 48444    35
```

```r
print(dim(symp19))
```

```
## [1] 60214    11
```

```r
print(dim(vax19))
```

```
## [1] 61204     8
```

```r
print(dim(data20))
```

```
## [1] 48901    35
```

```r
print(dim(symp20))
```

```
## [1] 60082    11
```

```r
print(dim(vax20))
```

```
## [1] 58622     8
```

```r
print(dim(data21))
```

```
## [1] 406001     35
```

```r
print(dim(symp21))
```

```
## [1] 542931     11
```

```r
print(dim(vax21))
```

```
## [1] 420981      8
```

```r
print(dim(confirmed))
```

```
## [1] 3193  555
```

```r
print(dim(population))
```

```
## [1] 3195    4
```

```r
print(dim(deaths))
```

```
## [1] 3193  555
```

Required Packages

- Readxl
- Tidyverse
- Stringr
- Dplyr
- Broom

- Scales
- Coefplot
- GGally
- QuantPsyc

Plots and Table Needs

- Histograms - Look for normality
- Scatterplots - Identify relationships
- Residual plots - Look for outliers
- Density plots - Observe distributions
- Box plots - Look for outliers
- Tables:
  - Vaccine types
  - Variables used
  - Common Symptoms
  - Reactions by state
  - Prior Allergies
  - Administered by variable explanation

Questions for future steps

I want to do research into the symptoms, but I am unsure of how far I will get due to my limitations in text data mining. Improve use of categorical variables in regression. Best way to handle missing data.

Reference:

VAERS Home. VAERS. (n.d.). https://vaers.hhs.gov/faq.html.