Ch3 in ROS, Probability and Confounding

Prof. Kapitula

9/16/2021

Israeli Covid Data

We will work with some data on Covid-19 Vaccine Efficacy, this is from https://www.covid-datascience.com/post/israeli-data-how-can-efficacy-vs-severe-disease-be-strong-when-60-of-hospitalized-are-vaccinated

```
library(tidyverse)
```

```
agegroup <- c("Under 50","Under 50","50 or older","50 or older","Total","Total")
vaxstatus <- c("Not Vax", "Fully Vax","Not Vax","Fully Vax", "Not Vax", "Fully Vax")
total <-c(1116834,3501118,186078 ,2133516,1302912,5634634)
severe<- c(43,11, 171,  290, 214,301)
delta <- tibble(agegroup,vaxstatus,total,severe)
rm(agegroup,vaxstatus,total,severe) #once the vectors are in a tibble delete them
delta
```

```
## # A tibble: 6 x 4
##   agegroup    vaxstatus   total severe
##   <chr>       <chr>       <dbl>  <dbl>
## 1 Under 50    Not Vax   1116834     43
## 2 Under 50    Fully Vax 3501118     11
## 3 50 or older Not Vax    186078    171
## 4 50 or older Fully Vax 2133516    290
## 5 Total       Not Vax   1302912    214
## 6 Total       Fully Vax 5634634    301
```

Claim: Almost 60% of those hospitalized or dead from Covid-19 in Israel as of August 15, 2021 were vaccinated.

Where did not that number come from? Use R as a calculator below:

```
301/(214+301)
```

```
## [1] 0.584466
```

To quote the original article:

> From many, I have seen this statistic used as evidence to support a narrative suggesting vaccines don't work or have lost their effectiveness vs. severe disease, and I have seen other articles quote this type of figure as further evidence for the reduction of effectiveness of the vaccines in trying to justify 3rd shot boosters.
>
> However, while these numbers are true, to quote them as evidence for low vaccine effectiveness is wrong and misleading. Sometimes, with observational data there is confounding of multiple factors that can make it easy to misinterpret simple percentages like this, and the current vaccination situation in Israel brings a perfect storm of confounding factors that lead to confusion if not thought through carefully.

In particular, the key factors here that contribute to this confusion are:

1. High vaccination rates in the country (nearly 80% of all residents >12yr)
2. Age disparity in vaccinations, including

- Nearly all older people being vaccinated (>90% of residents >50yr) and
- The vast majority of unvaccinated being younger people (>85% of unvaccinated <50yr)
- Older people are orders of magnitude more likely to be hospitalized with a respiratory virus than young people (residents >50yr are >20x more likely to have hospitalized serious infections than residents <50yr, and residents 90+ are >1600x more likely to have hospitalized serious infections than residents 12-15yr)

Data Wrangle

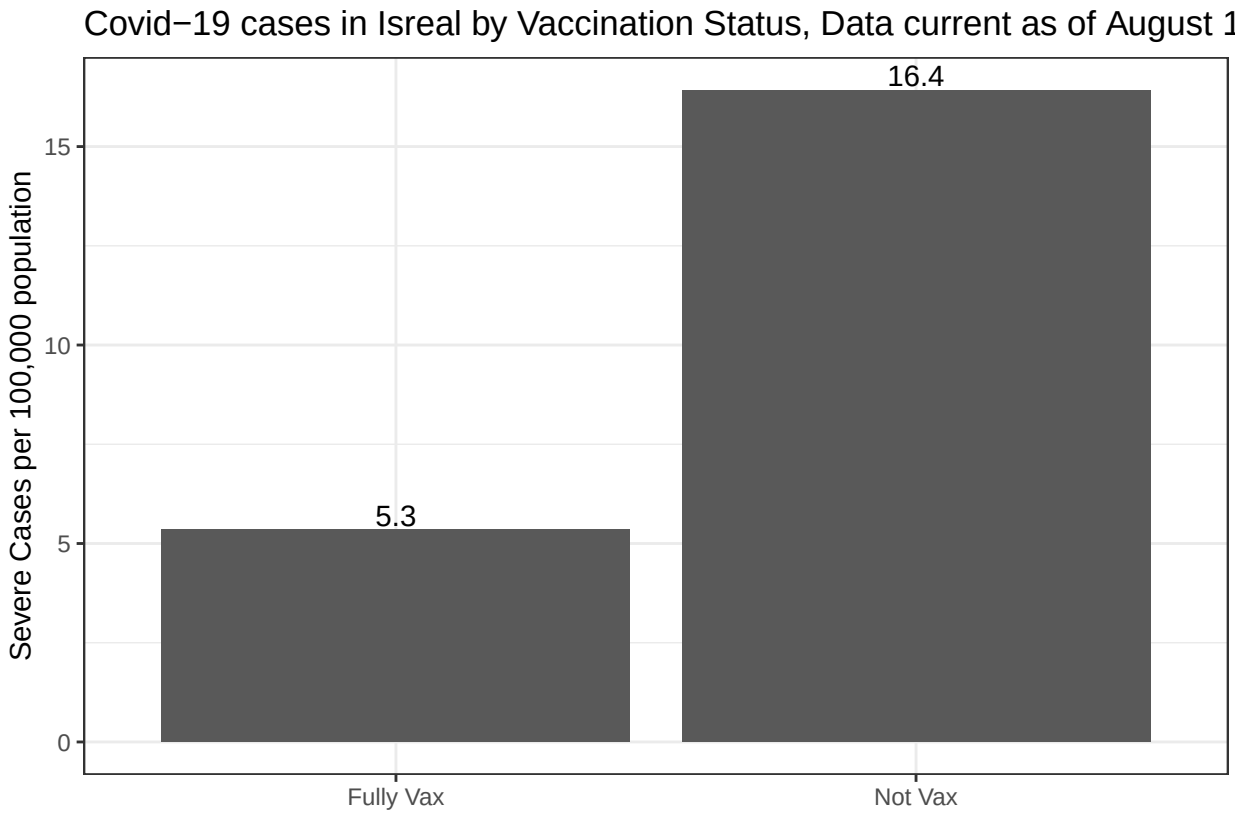
```
delta <-
  delta %>%
  mutate(severeper100k=(severe/total)*100000)

totals <- delta %>%
  filter(agegroup=="Total")

byage <- delta %>%
  filter(agegroup!="Total")
```

What is the code above doing? Making two different data sets, one called totals that does not breakdown by age, and one called byage that groups by age.

Why do we use filter instead of select to make the two new data sets? Because we are selecting cases (filtering rows) not selecting variables.

Plot ignoring age

```
ggplot(totals,aes(vaxstatus,severeper100k)) +
  geom_col() +
  labs(x="", y="Severe Cases per 100,000 population",
       title="Covid-19 cases in Isreal by Vaccination Status, Data current as of August 15, 2021") +ther
  geom_text(aes(label = round(severeper100k,1)), vjust = -.2)
```



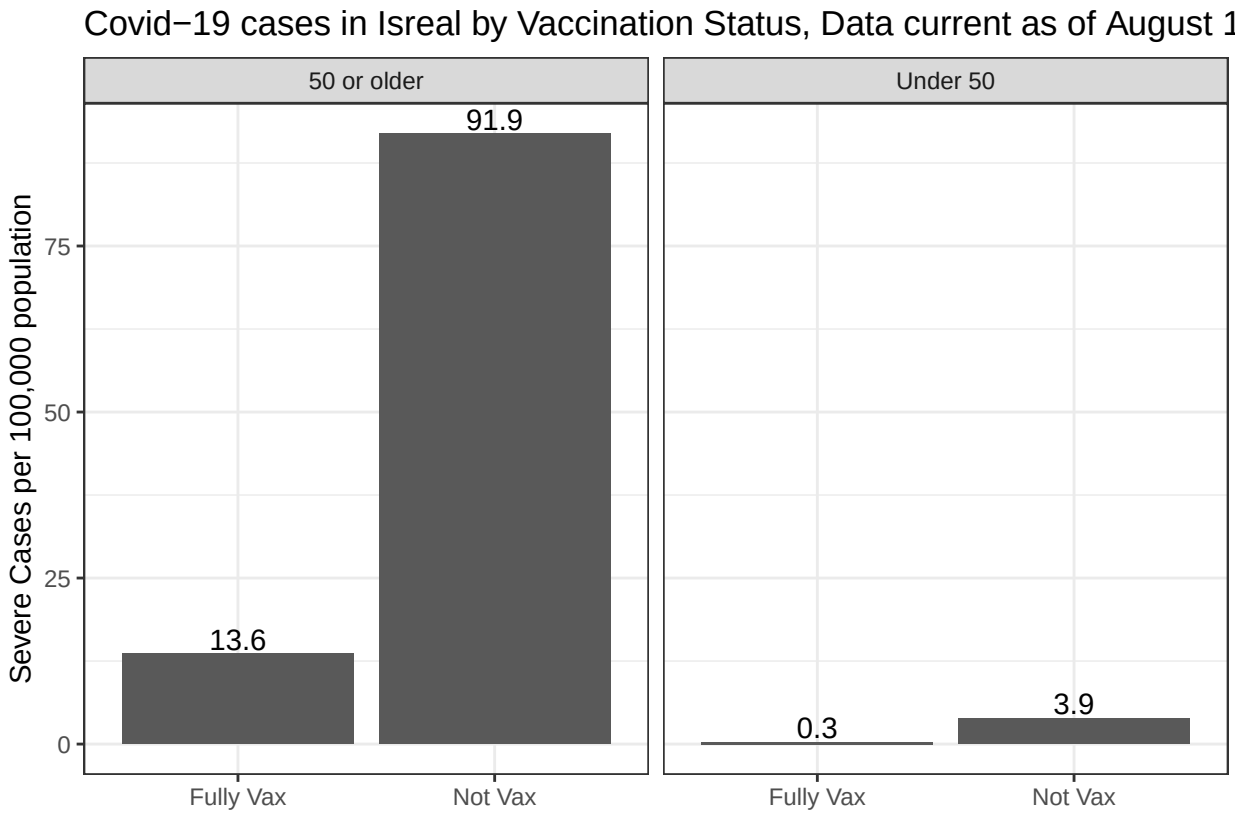
What is the vaccine efficacy against severe covid-19? To get this value divide the difference in rates by the rate for the unvaccinated. You can just use R as a calculator to do this, note that we can reference original values and variables.

```
(totals$severeper100k[1]-totals$severeper100k[2])/totals$severeper100k[1]
```

```
## [1] 0.6747614
```

Next, copy the code above. Recreate the graphic faceted by age group. Look at https://ggplot2.tidyverse.org/reference/ . Remember you will need to change the data you are using.

```
ggplot(byage,aes(vaxstatus,severeper100k)) +
  geom_col() +
  labs(x="", y="Severe Cases per 100,000 population",
       title="Covid-19 cases in Isreal by Vaccination Status, Data current as of August 15, 2021") +ther
  geom_text(aes(label = round(severeper100k,1)), vjust = -.2)+
  facet_grid(cols = vars(agegroup))
```



Calculate the efficacy within each group: (hint, use byage, copy code above to calculate efficacy, you will have two separate values)

```
byage
```

```
## # A tibble: 4 x 5
##   agegroup    vaxstatus   total severe severeper100k
##   <chr>       <chr>       <dbl>  <dbl>         <dbl>
## 1 Under 50    Not Vax   1116834     43         3.85
## 2 Under 50    Fully Vax 3501118     11         0.314
## 3 50 or older Not Vax    186078    171        91.9
## 4 50 or older Fully Vax 2133516    290        13.6
```

```
(byage$severeper100k[1]-byage$severeper100k[2])/byage$severeper100k[1]
```

```
## [1] 0.918397
```

```
(byage$severeper100k[3]-byage$severeper100k[4])/byage$severeper100k[3]
```

```
## [1] 0.8520888
```

This is why we need to look at weighted averages and think about confounding.