DATA607 - Assignment Week 5
Tidying and Transforming Vaccination Data

Donald Butler

9/26/2021

Introduction

Read a csv of vaccination data in Israel taken from August 2021.

Load required R Libraries

```
library(tidyverse)
```

Import and Clean Data

Import raw vaccination data

```
infoFile = 'https://raw.githubusercontent.com/dab31415/DATA607/main/Homework/Assignment_4/israeli_vacci
raw_vaccination <-read.table(file = infoFile,
                             sep = ',',
                             strip.white = TRUE,
                             fill = TRUE)

head(raw_vaccination,7)
```

```
##      V1         V2           V3                             V4                  V5
## 1 Age Population %                            Severe Cases
## 2       Not Vax\n% Fully Vax\n% Not Vax\nper 100K\n\n\np Fully Vax\nper 100K
## 3 <50    1,116,834    3,501,118                             43                  11
## 4              23.3%        73.0%
## 5 >50      186,078    2,133,516                            171                 290
## 6               7.9%        90.4%
## 7
##                  V6
## 1          Efficacy
## 2 vs. severe disease
## 3
## 4
## 5
## 6
## 7
```

Extract Required Data

The data we are looking to extract from the csv file is in columns 1-5 on rows 3 and 5, and columns 2-3 on rows 4 and 6.

```
vaccination <- raw_vaccination[c(3,5),c(1:5)] %>%
  cbind(raw_vaccination[c(4,6),c(2:3)])

names(vaccination) <- c('ageGroup','not_vaccinated','fully_vaccinated','not_vaccinated_severe','fully_va

(vaccination <- as_tibble(vaccination))
```

```
## # A tibble: 2 x 7
##   ageGroup not_vaccinated fully_vaccinated not_vaccinated_se~ fully_vaccinated_~
##   <chr>    <chr>          <chr>            <chr>              <chr>
## 1 <50      1,116,834      3,501,118        43                 11
## 2 >50      186,078        2,133,516        171                290
## # ... with 2 more variables: not_vaccinated_percent <chr>,
## #   fully_vaccinated_percent <chr>
```

Convert column data types

```
vaccination <- vaccination %>%
  transmute(
    ageGroup,
    not_vaccinated = as.integer(str_replace_all(not_vaccinated,',','')),
    fully_vaccinated = as.integer(str_replace_all(fully_vaccinated,',','')),
    not_vaccinated_severe = as.integer(not_vaccinated_severe),
    fully_vaccinated_severe = as.integer(fully_vaccinated_severe),
    not_vaccinated_percent = as.numeric(str_replace(not_vaccinated_percent,'%','')) / 100,
    fully_vaccinated_percent = as.numeric(str_replace(fully_vaccinated_percent,'%','')) / 100
  )
```

Calcualate Partially Vaccinated Populations

Israel is using the Pfizer Covid vaccine which requires two doses before being considered fully vaccinated. For each age group, the percentage of not vaccinated and fully vaccinated can be used to calculate the total population in the group and the number of people that are partially vaccinated.

```
vaccination <- vaccination %>%
  mutate(partially_vaccinated = round((not_vaccinated + fully_vaccinated) / (not_vaccinated_percent + fu
```

Tidy population and severe cases by vaccination status

```
a <- vaccination %>%
  select(ageGroup, not_vaccinated, partially_vaccinated, fully_vaccinated) %>%
  pivot_longer(cols=2:4, names_to = 'vaccination_status',values_to = 'population')
```

```
b <- vaccination %>%
  select(ageGroup,not_vaccinated = not_vaccinated_severe,fully_vaccinated = fully_vaccinated_severe) %>%
  pivot_longer(cols=2:3, names_to = 'vaccination_status', values_to = 'severe_cases')

tidy_vaccinations <- a %>%
  left_join(b, by = c('ageGroup' = 'ageGroup', 'vaccination_status' = 'vaccination_status'))

tidy_vaccinations
```

```
## # A tibble: 6 x 4
##   ageGroup vaccination_status   population severe_cases
##   <chr>    <chr>                     <dbl>        <int>
## 1 <50      not_vaccinated          1116834           43
## 2 <50      partially_vaccinated     177429           NA
## 3 <50      fully_vaccinated        3501118           11
## 4 >50      not_vaccinated           186078          171
## 5 >50      partially_vaccinated      40115           NA
## 6 >50      fully_vaccinated        2133516          290
```

Question 1

Do you have enough information to calculate the total population? What does this total population represent?

```
(total_population = sum(tidy_vaccinations$population))
```

```
## [1] 7155090
```

I initially thought that there was enough information to calculate the total population after adding in the number of partially vaccinated individuals for each age group. Based on the dataset, I've calculated the population of Israel as 7.2 million people, however looking online I found the total population is about 8.8 million. This seems to indicate that we have not accounted for about 1.6 million people.

Without knowing the specifics of the dataset, I suspect the difference in population counts may indicate that they have excluded children under the age of 12 which are currently unable to get vaccinated.

Question 2

Calculate the Efficacy vs Disease using the formula 1 - ((% fully vaccinated severe cases per 100k) / (% not vaccinated severe cases per 100k)).

```
(U50Efficacy <- 1 - (11/43))
```

```
## [1] 0.744186
```

```
(O50Efficacy <- 1 - (290/171))
```

```
## [1] -0.6959064
```

In the over 50 age group the vaccine appears to not be effective at preventing severe cases, but this is due to the increase in cases of the delta variant, the low number of non-vaccinated people in the age group, and the amount of time since completion of their second dose.

In the under 50 age group, the vaccine efficacy is high indicating that it is better than being unvaccinated.

Question 3

From your calculation of efficacy vs. disease, are you able to compare the rate of severe cases in unvaccinated individuals to that in vaccinated individuals?

I'm not clear what is being asked here. I would have thought that the rates of severe cases between the two groups is in the severe cases per 100K statistics.