Emmalyn Campau
STATS485

# Investigating the Efficacy of the COVID-19 Vaccine

***Introduction***

In 2021, mRNA vaccines for COVID-19 were approved for emergency use in the United States. This approval came after clinical trials found these vaccines to be over 90% effective, but these trials preceded the emergence of the highly contagious Delta variant. The surge of this variant raised concerns that the vaccines might not be as effective against new mutations. This study looked at a table of data collected from Israel's government-owned COVID-19 dashboard to confront these concerns. The Agresti-Coull confidence interval was applied to this data to investigate the effectiveness of the Pfizer-BioNTech and Moderna vaccines in preventing severe cases after the rise of the Delta variant. The 95% confidence interval for the severe COVID rates in vaccinated and unvaccinated patients were taken separately for each age group. Next, the confidence intervals for the vaccine efficacy within each age group were calculated. The final calculation was the age-standardized vaccine efficacy for all individuals age 12 and older, which found the vaccines to be effective against the Delta variant.

***Methods***

The data used in this statistical analysis was originally collected from Israel's COVID-19 dashboard on August 15, 2021 and contains unvaccinated and vaccinated subpopulations, which have been further stratified into seven age groups. The unvaccinated and vaccinated subpopulation totals reflect a census of Israeli citizens and permanent residents. The number of severe COVID-19 cases for both unvaccinated and vaccinated people in each age group was also recorded, where severe cases were defined by hospitalizations. The total number of samples in this dataset was 6,937,542 people.

After the data was obtained, a function that takes in any $n$ and $\hat{p}$ values and then calculates the 95% confidence interval using the Agresti-Coull interval as defined in “Interval Estimation for a Binomial Proportion” (Brown et al. 2001) was constructed. The Agresti-Coull interval was chosen over the Wald interval because it retains simplicity while improving coverage performance, especially when p is near 0 or 1. This means the probability that the confidence interval will include the true value is higher than in the Wald interval, which is often used despite having bad coverage. Like many statistical methods, the Agresti-Coull interval makes some assumptions about the data being used, including that the data is binomial, has a large sample size, and is randomly sampled and independent. The number of cases used in this study’s dataset was binomial because each person either was or was not hospitalized, and each of these people had the same probability of infection.

The Agresti-Coull function was then used to estimate the 95% confidence intervals for the underlying rates of severe disease in the vaccinated ($p_a^{(v)}$) and unvaccinated ($p_a^{(u)}$) subpopulations for each age group, $a$. For each of these age groups, the observed proportion of people with the vaccine that were hospitalized was passed into the function as $\hat{p}$ and the number of vaccinated people in the group was passed in as $n$. The results were the confidence intervals for the estimated rate of severe disease in vaccinated individuals for each group. The same process was repeated with the proportions of unvaccinated hospitalizations and the number of unvaccinated individuals.

Then, the point estimates and corresponding confidence intervals for the vaccine efficacy for each age group were calculated. The vaccine efficacy is defined by the equation $1 - p_a^{(v)}/p_a^{(u)}$ for each age group, $a$, and presupposes the existence of vaccinated and unvaccinated subpopulations. In this case, the vaccine efficacy is a measure that compares the percentage of individuals with the vaccine who developed severe illness to the proportion of people without the vaccine that developed severe illness. For example, a vaccine efficacy of 0.80 indicates that those who were vaccinated were at a 80% lower risk of developing severe illness than those who were unvaccinated.

The point estimates of the vaccine efficacy for each age group were calculated by plugging in the values from the dataset into the formula $1 - p_a^{(v)}/p_a^{(u)}$. The proportion of severe disease cases among vaccinated people in an age group was used as $p_a^{(v)}$, and the proportion of severe disease cases among unvaccinated people in the same age group was used as $p_a^{(u)}$. Afterwards, the confidence intervals were calculated by finding the largest and smallest vaccine efficacies that are compatible with the confidence intervals reported in the first step of the analysis. This meant that the lower bound of the vaccine efficacy confidence interval was determined by $1 - \frac{U^v}{L^u}$ and the upper bound was determined by $1 - \frac{L^v}{U^u}$. In these equations, $U^v$ is the upper bound and $L^v$ is the lower bound of the confidence interval for $\widehat{p}^{(v)}$, and $U^u$ is the upper bound and $L^u$ is the lower bound of the confidence interval for $\widehat{p}^{(u)}$. This calculation was repeated for each of the separate age groups. The confidence intervals for the vaccine efficacy were at the 90% confidence level because they were based on the intersection of two 95% confidence intervals, meaning the new confidence level is $0.95^2$.

The final step of the analysis was to calculate the age-standardized vaccine efficacy. To do this, the estimated vaccine efficacy of each age group, found previously, was multiplied by the proportion of the total population in that age group (Mathieu 2023). These seven percentages, one for each age group, were added together to estimate the total age-standardized vaccine efficacy.

***Results***

The confidence intervals for underlying rates of severe disease in the vaccinated and unvaccinated subpopulations for each age group are shown in Table 1 below. The table contains columns for the lower and upper limits of each confidence interval, and the rows are separated by age range.

| Age Group | Unvaccinated CI Lower Bound | Unvaccinated CI Upper Bound | Vaccinated CI Lower Bound | Vaccinated CI Upper Bound |
|---|---|---|---|---|
| 12-39 | 1.23 | 3.07 | 0.00155 | .303 |
| 40-49 | 10.9 | 24.7 | 0.479 | 1.88 |
| 50-59 | 28.6 | 56.4 | 1.92 | 4.48 |
| 60-69 | 58.0 | 101.0 | 6.72 | 11.3 |
| 70-79 | 138.0 | 261.0 | 16.1 | 24.3 |
| 80-89 | 178.0 | 357.0 | 39.3 | 58.3 |
| 90+ | 308.0 | 835.0 | 24.0 | 61.5 |
| all 12+ | 14.4 | 18.8 | 4.77 | 5.98 |

**Table 1** | The table contains the confidence intervals for underlying rates of severe disease in the vaccinated and unvaccinated subpopulations in each age group. The bottom row shows the statistics for the entire sample of people, without excluding ages. Results are reported as the estimated number of cases per 100,000 people.

All of the percentages reported in the confidence intervals are small decimals, with $p$ close to 0. However, it can be noted that the estimated bounds of the confidence intervals for the unvaccinated subpopulation are larger than the bounds of the confidence intervals for the vaccinated subpopulation by one to three decimal places. It also appears that the rate of disease generally increases with age, regardless of vaccination status.

The point estimates and estimated confidence intervals for the vaccine efficacy within each age range are reported in Table 2, where the columns are the point estimate and upper and lower bounds for the confidence interval, and the rows are once again separated by age range.

| Age Group | Vaccine Efficacy Point Estimate | Vaccine Efficacy Lower Bound | Vaccine Efficacy Upper Bound |
|---|---|---|---|
| 12-39 | 0.960 | 0.754 | 0.999 |
| 40-49 | 0.941 | 0.829 | 0.981 |
| 50-59 | 0.927 | 0.843 | 0.966 |
| 60-69 | 0.886 | 0.805 | 0.934 |
| 70-79 | 0.896 | 0.824 | 0.938 |
| 80-89 | 0.810 | 0.672 | 0.890 |
| 90+ | 0.924 | 0.800 | 0.971 |
| all 12+ | 0.675 | 0.584 | 0.746 |

**Table 2** | The rows in this table are separated by age group, with the bottom row showing the results for the entire sample. The columns included are the point estimate for vaccine efficacy and the lower and upper bounds of the estimated vaccine efficacy confidence interval.

The results show that the estimated vaccine efficacy is quite high in younger age groups, with the age range 12-39 having a point estimate of 0.96. However, as the age increases, the point estimate, lower bound, and upper bound of the vaccine efficacy typically decrease. Interestingly, in the oldest age group, those 90 years of age or older, efficacy spikes again, and the point estimate is 0.924. Not including the point estimate for the total sample, the lowest that the point estimate drops is in those aged 80-89 with an estimate of 0.810, which is 0.15 lower than the efficacy of the youngest age group.

The final calculation was the estimated age-standardized vaccine efficacy. Before age-standardization, the estimated vaccine efficacy was 0.675, which is also shown in Table 2. After standardizing the age, the estimated vaccine efficacy rose to 0.944.

***Discussion***

The results presented showed that, for all age groups, the rate of severe disease was higher in the unvaccinated subpopulation than in the vaccinated subpopulation. It was also shown that the vaccine efficacy usually decreases with age, but standardizing the vaccine efficacy by age group shows that the vaccine is still very effective against the Delta variant.

These calculations utilized the Agresti-Coull confidence interval, which has many advantages over the well-known Wald interval that is taught in most introductory statistics courses. One reason the Agresti-Coull interval was chosen over alternatives is its ability to perform well for large sample sizes, which fits this dataset's sample size of 6,937,542. The next reason was that the Agresti-Coull has a more conservative approach for $p$ very close to 0 or 1, meaning it sacrifices specificity in order to increase the chance the true value will be included in the estimated interval, and the proportions of individuals with severe disease were small decimals that were close to 0. This leads us to a potential limitation of this interval, which is that it has a longer length and therefore is not quite as specific. Because the interval retains simplicity in its formula compared to alternatives, it was determined that a less preferable length was a tradeoff worth making. All these properties of the Agresti-Coull interval arguably fit the qualities of the data and purpose of the report.

Along with choosing a fitting confidence interval method, it was also important to standardize the vaccine efficacy by age to avoid reporting a misleading percentage. When calculating the vaccine efficacy without standardization, the result was an efficacy of only 0.675, which is a vastly different result than the age-standardized percentage of 0.944. This difference is due to a variety of confounding factors and the large disparity in vaccinated populations. A larger percentage of older people are vaccinated than the younger age groups included in the data, with over 90% of those older than 50 years old having the vaccine. Another aspect to be taken into consideration is that older people are much more likely to be hospitalized after developing respiratory diseases than younger people are, regardless of vaccination status. This explains the observed increase in severe disease and decreased vaccine efficacy as age increases. Finding the vaccine efficacy of the total population without first standardizing by age ignores these considerations and is widely misleading.

The standardized vaccine efficacy suggests that these vaccines are highly effective against the Delta variant. In fact, all of the findings highlight the effectiveness of the vaccine and show that the unvaccinated subpopulations were hospitalized in larger proportions than the vaccinated people. Going forward, it will be important to continue collecting COVID-19 data and investigating the effectiveness of the vaccines to spread public awareness and avoid unsubstantiated distrust in the vaccine.

References

Brown, Lawrence D., et al. “Interval Estimation for a Binomial Proportion”. *Statistical Science,* vol. 16, no. 2, 2001, pp 101-133

Mathieu, Edouard. “How does age standardization make health metrics comparable?”. *Our World in Data.* April 04, 2023. https://ourworldindata.org/age-standardization#:~:text=Age%20standardization%20is%20a%20statistical,differences%20in%20their%20age%20structure.

Morris, Jeffrey. “Israeli data: How can efficacy vs. severe disease be strong when 60% of hospitalized are vaccinated?”. *COVID-19 Data Science*. October 21, 2021. https://www.covid-datascience.com/post/israeli-data-how-can-efficacy-vs-severe-disease-be-strong-when-60-of-hospitalized-are-vaccinated

“Vaccine efficacy, effectiveness and protection”. *World Health Organization*. July 14, 2021 https://www.who.int/news-room/feature-stories/detail/vaccine-efficacy-effectiveness-and-protection

# Unit 1 Paper Version 2

Emmalyn Campau

2023-09-24

This file contains code that analyzes a dataset from Israels' COVID-19 dashboard. The goal of the project is to determine rates of severe disease and to estimate vaccine efficacy. The 95% confidence interval for the estimated rates of disease in both vaccinated and unvaccinated subpopulations was calculated using the Agresti-Coull method for each age group and for the entire sample. Next, the estimated vaccine efficacies for each of these age groups were calculated. Finally, I calculated the age-standardized vaccine efficacy for the entire sample.

```
# Read in the data for vaccines and COVID in Israel
# Store in dataset titled 'data'
data <-
read.csv("https://dept.stat.lsa.umich.edu/~bbh/s485/data/israeli_covid_2021_0
8_15.csv")
```

Part 1: In this section of code, the 95% confidence interval for the underlying rates of severe disease in the vaccinated (pa(v)) and unvaccinated (pa(u)) subpopulations was calculated for each age group.

```
# For each age group a, a 95% confidence interval for the underlying rates of
severe disease in the vaccinated (pa(v)) and unvaccinated (pa(u))
subpopulations

# function - define the Agresti-Coull Interval (95% interval)
ac_CI <- function(n, phat) {
  k = qnorm(0.975) # same thing as qnorm(1 - (1 - confidence level) / 2)
  #Agresti-Coull
  X_tilde = ((phat*n) + k^2/2)
  n_tilde = n + k^2
  p_tilde = X_tilde / n_tilde
  q_tilde = 1 - p_tilde

  upper_bound <- p_tilde + (k*sqrt(p_tilde*q_tilde)/sqrt(n_tilde))
  lower_bound <- p_tilde - (k*sqrt(p_tilde*q_tilde)/sqrt(n_tilde))
  interval <- c(lower_bound, upper_bound)
  interval
}

# sanity check for the Agresti Coull funciton
# function output
```

```r
test_function <- ac_CI(2573902, 2/2573903)
test_function

## [1] 1.547852e-08 3.031042e-06

# actual
actual_lower <- ((2 +(1.96^2)/2) / (2573903 + (1.96^2))) - (1.96*sqrt(((2
+(1.96^2)/2)) /
                  (2573903 + (1.96^2))*(1 - (((2 +(1.96^2)/2)) / (2573903 +
(1.96^2)))))/sqrt(2573903 + (1.96^2)))
actual_lower

## [1] 1.546481e-08

actual_upper <- ((2 +(1.96^2)/2) / (2573903 + (1.96^2))) + (1.96*sqrt(((2
+(1.96^2)/2)) /
                  (2573903 + (1.96^2))*(1 - (((2 +(1.96^2)/2)) / (2573903 +
(1.96^2)))))/sqrt(2573903 + (1.96^2)))
actual_upper

## [1] 3.03111e-06

# Age group 12-39 [1]
#vaccinated
phat_vax_1 <- (2/2573903)
vax_interval_1 <- ac_CI(2573902, phat_vax_1)
vax_interval_1

## [1] 1.547852e-08 3.031042e-06

#unvaccinated
phat_unvax_1 <- (19/971478)
unvax_interval_1 <- ac_CI(971478, phat_unvax_1)
unvax_interval_1

## [1] 1.230709e-05 3.076264e-05

# Age group 40-49 [2]
#vaccinated
phat_vax_2 <- (9/927214)
vax_interval_2 <- ac_CI(927214, phat_vax_2)
vax_interval_2

## [1] 4.792582e-06 1.876333e-05

#unvaccinated
phat_unvax_2 <- (24/145355)
```

```r
unvax_interval_2 <- ac_CI(145355, phat_unvax_2)
unvax_interval_2
```

```
## [1] 0.0001096801 0.0002469646
```

```r
# Age group 50-59 [3]
#vaccinated
phat_vax_3 <- (22/747949)
vax_interval_3 <- ac_CI(747949, phat_vax_3)
vax_interval_3
```

```
## [1] 1.916555e-05 4.479765e-05
```

```r
#unvaccinated
phat_unvax_3 <- (34/84545)
unvax_interval_3 <- ac_CI(84545, phat_unvax_3)
unvax_interval_3
```

```
## [1] 0.0002859459 0.0005637577
```

```r
# age group 60-69 [4]
#vaccinated
phat_vax_4 <- (58/665717)
vax_interval_4 <- ac_CI(665717, phat_vax_4)
vax_interval_4
```

```
## [1] 6.721981e-05 1.127978e-04
```

```r
#unvaccinated
phat_unvax_4 <- (50/65205)
unvax_interval_4 <- ac_CI(65205, phat_unvax_4)
unvax_interval_4
```

```
## [1] 0.0005797316 0.0010127128
```

```r
# age group 70-79 [5]
#vaccinated
phat_vax_5 <- (92/464336)
vax_interval_5 <- ac_CI(464336, phat_vax_5)
vax_interval_5
```

```
## [1] 0.0001613648 0.0002431696
```

```r
#unvaccinated
phat_unvax_5 <- (39/20512)
unvax_interval_5 <- ac_CI(20512, phat_unvax_5)
unvax_interval_5
```

```
## [1] 0.001384076 0.002605108
```

```r
# age group 80-89 [6]
#vaccinated
phat_vax_6 <- (100/208911)
vax_interval_6 <- ac_CI(208911, phat_vax_6)
vax_interval_6
```

```
## [1] 0.0003931678 0.0005825478
```

```r
#unvaccinated
phat_unvax_6 <- (32/12683)
unvax_interval_6 <- ac_CI(12683, phat_unvax_6)
unvax_interval_6
```

```
## [1] 0.001775136 0.003572251
```

```r
# age group 90+ [7]
#vaccinated
phat_vax_7 <- (18/46602)
vax_interval_7 <- ac_CI(46602, phat_vax_7)
vax_interval_7
```

```
## [1] 0.0002397717 0.0006150881
```

```r
#unvaccinated
phat_unvax_7 <- (16/3132)
unvax_interval_7 <- ac_CI(3132, phat_unvax_7)
unvax_interval_7
```

```
## [1] 0.003076488 0.008353126
```

```r
# age group all 12+ [8]
#vaccinated
phat_vax_8 <- (301/5634632)
vax_interval_8 <- ac_CI(5634632, phat_vax_8)
vax_interval_8
```

```
## [1] 4.770658e-05 5.981438e-05
```

```r
#unvaccinated
phat_unvax_8 <- (214/1302910)
unvax_interval_8 <- ac_CI(1302910, phat_unvax_8)
unvax_interval_8
```

```
## [1] 0.0001436188 0.0001878240
```

```r
#display results in a table
matrix_data = matrix(c(vax_interval_1, unvax_interval_1, vax_interval_2,
                       unvax_interval_2, vax_interval_3, unvax_interval_3,
                       vax_interval_4, unvax_interval_4, vax_interval_5,
                       unvax_interval_5, vax_interval_6, unvax_interval_6,
                       vax_interval_7, unvax_interval_7, vax_interval_8,
                       unvax_interval_8), ncol=4, byrow=TRUE)
#matrix_data
# specify the column names and row names of matrix
colnames(matrix_data) = c('Vax Lower Limit','Vax Upper Limit', 'Unvax Lower
Limit', 'Unvax Upper Limit')
rownames(matrix_data) <- c('12-39','40-49','50-59','60-69', '70-79', '80-89',
'90+', 'all 12+')

# assign to table
final=as.table(matrix_data)

# display
final

##           Vax Lower Limit Vax Upper Limit Unvax Lower Limit Unvax Upper
Limit
## 12-39    1.547852e-08     3.031042e-06     1.230709e-05
3.076264e-05
## 40-49    4.792582e-06     1.876333e-05     1.096801e-04
2.469646e-04
## 50-59    1.916555e-05     4.479765e-05     2.859459e-04
5.637577e-04
## 60-69    6.721981e-05     1.127978e-04     5.797316e-04
1.012713e-03
## 70-79    1.613648e-04     2.431696e-04     1.384076e-03
2.605108e-03
## 80-89    3.931678e-04     5.825478e-04     1.775136e-03
3.572251e-03
## 90+      2.397717e-04     6.150881e-04     3.076488e-03
8.353126e-03
## all 12+  4.770658e-05     5.981438e-05     1.436188e-04
1.878240e-04
```

Part 2: In this section of code, the vaccine efficacy was calculated for each age group and for the total population.

```r
# Equation for vaccine efficacy
efficacy <- function(p_vax, p_unvax) {
  true_values = 1 - (p_vax/p_unvax)
```

```r
  true_values
}
# For point estimates, run function efficacy with phats from part 1

# Age group 12-39 [1]
point_val1 <- efficacy(phat_vax_1, phat_unvax_1)
point_val1 <- signif(point_val1, 3)
point_val1
```

```
## [1] 0.96
```

```r
# Find the lower bound using the largest value of pv/pu
vax_interval_1[1]
```

```
## [1] 1.547852e-08
```

```r
vax_interval_1[2]
```

```
## [1] 3.031042e-06
```

```r
estimate_lower1 <- efficacy(vax_interval_1[2], unvax_interval_1[1])
estimate_lower1 <- signif(estimate_lower1, 3)
estimate_lower1
```

```
## [1] 0.754
```

```r
# Find the upper bound using the smallest value of pv/pu
estimate_upper1 <- efficacy(vax_interval_1[1], unvax_interval_1[2])
estimate_upper1 <- signif(estimate_upper1, 3)
estimate_upper1
```

```
## [1] 0.999
```

```r
# Age group 40-49 [2]
# point estimate for ages 40-49
point_val2 <- efficacy(phat_vax_2, phat_unvax_2)
point_val2 <- signif(point_val2, 3)
point_val2
```

```
## [1] 0.941
```

```r
# confidence interval
# Find the lower bound using the largest value of pv/pu
vax_interval_2[2]
```

```
## [1] 1.876333e-05
```

```r
unvax_interval_2[1]
```

```r
## [1] 0.0001096801

estimate_lower2 <- 1 - (vax_interval_2[2]/unvax_interval_2[1])
estimate_lower2 <- signif(estimate_lower2, 3)
estimate_lower2

## [1] 0.829

estimate_upper2 <- efficacy(vax_interval_2[1], unvax_interval_2[2])
estimate_upper2 <- signif(estimate_upper2, 3)
estimate_upper2

## [1] 0.981

# Age group 50-59 [3]
# point estimate for ages 50-59
point_val3 <- efficacy(phat_vax_3, phat_unvax_3)
point_val3 <- signif(point_val3, 3)
point_val3

## [1] 0.927

# confidence interval
estimate_lower3 <- efficacy(vax_interval_3[2], unvax_interval_3[1])
estimate_lower3 <- signif(estimate_lower3, 3)
estimate_lower3

## [1] 0.843

estimate_upper3 <- efficacy(vax_interval_3[1], unvax_interval_3[2])
estimate_upper3 <- signif(estimate_upper3, 3)
estimate_upper3

## [1] 0.966

# Age group 60-69 [4]
# point estimate for ages 60-69
point_val4 <- efficacy(phat_vax_4, phat_unvax_4)
point_val4 <- signif(point_val4, 3)
point_val4

## [1] 0.886

# confidence interval
estimate_lower4 <- efficacy(vax_interval_4[2], unvax_interval_4[1])
estimate_lower4 <- signif(estimate_lower4, 3)
estimate_lower4
```

```
## [1] 0.805

estimate_upper4 <- efficacy(vax_interval_4[1], unvax_interval_4[2])
estimate_upper4 <- signif(estimate_upper4, 3)
estimate_upper4

## [1] 0.934

# Age group 70-79 [5]
# point estimate for ages 70-79
point_val5 <- efficacy(phat_vax_5, phat_unvax_5)
point_val5 <- signif(point_val5, 3)
point_val5

## [1] 0.896

# confidence interval
estimate_lower5 <- efficacy(vax_interval_5[2], unvax_interval_5[1])
estimate_lower5 <- signif(estimate_lower5, 3)
estimate_lower5

## [1] 0.824

estimate_upper5 <- efficacy(vax_interval_5[1], unvax_interval_5[2])
estimate_upper5 <- signif(estimate_upper5, 3)
estimate_upper5

## [1] 0.938

# Age group 80-89 [6]
# point estimate for ages 80-89
point_val6 <- efficacy(phat_vax_6, phat_unvax_6)
point_val6 <- signif(point_val6, 3)
point_val6

## [1] 0.81

# confidence interval
estimate_lower6 <- efficacy(vax_interval_6[2], unvax_interval_6[1])
estimate_lower6 <- signif(estimate_lower6, 3)
estimate_lower6

## [1] 0.672

estimate_upper6 <- efficacy(vax_interval_6[1], unvax_interval_6[2])
estimate_upper <- signif(estimate_upper6, 3)
estimate_upper6
```

```
## [1] 0.8899384

# Age group 90+ [7]
# point estimate for ages 90+
point_val7 <- efficacy(phat_vax_7, phat_unvax_7)
point_val7 <- signif(point_val7, 3)
point_val7

## [1] 0.924

# confidence interval
estimate_lower7 <- efficacy(vax_interval_7[2], unvax_interval_7[1])
estimate_lower7 <- signif(estimate_lower7, 3)
estimate_lower7

## [1] 0.8

estimate_upper7 <- efficacy(vax_interval_7[1], unvax_interval_7[2])
estimate_upper7 <- signif(estimate_upper7, 3)
estimate_upper7

## [1] 0.971

# Age group all 12+ [8]
# point estimate for all ages 12+
point_val8 <- efficacy(phat_vax_8, phat_unvax_8)
point_val8 <- signif(point_val8, 3)
point_val8

## [1] 0.675

# confidence interval
estimate_lower8 <- efficacy(vax_interval_8[2], unvax_interval_8[1])
estimate_lower8 <- signif(estimate_lower8, 3)
estimate_lower8

## [1] 0.584

estimate_upper8 <- efficacy(vax_interval_8[1], unvax_interval_8[2])
estimate_upper8 <- signif(estimate_upper8, 3)
estimate_upper8

## [1] 0.746

# Organize results into table
matrix_data_efficacy = matrix(c(point_val1, estimate_lower1, estimate_upper1,
                               point_val2, estimate_lower2, estimate_upper2,
                               point_val3, estimate_lower3, estimate_upper3,
```

```
                                 point_val4, estimate_lower4, estimate_upper4,
                                 point_val5, estimate_lower5, estimate_upper5,
                                 point_val6, estimate_lower6, estimate_upper6,
                                  point_val7, estimate_lower7, estimate_upper7,
                                 point_val8, estimate_lower8, estimate_upper8),
                                 ncol=3, byrow=TRUE)
#matrix_data
# specify the column names and row names of matrix
colnames(matrix_data_efficacy) = c('Vaccine Efficacy Point Estimate',
                                   'Vaccine Efficacy Lower Limit',
                                   'Vaccine Efficacy Upper Limit')
rownames(matrix_data_efficacy) <- c('12-39','40-49','50-59','60-69', '70-79',
                                   '80-89', '90+', 'all 12+')

# assign to table
final=as.table(matrix_data_efficacy)

# display
final

##           Vaccine Efficacy Point Estimate Vaccine Efficacy Lower Limit
## 12-39                          0.9600000                    0.7540000
## 40-49                          0.9410000                    0.8290000
## 50-59                          0.9270000                    0.8430000
## 60-69                          0.8860000                    0.8050000
## 70-79                          0.8960000                    0.8240000
## 80-89                          0.8100000                    0.6720000
## 90+                               0.9240000                 0.8000000
## all 12+                        0.6750000                    0.5840000
##           Vaccine Efficacy Upper Limit
## 12-39                          0.9990000
## 40-49                          0.9810000
## 50-59                          0.9660000
## 60-69                          0.9340000
## 70-79                          0.9380000
## 80-89                          0.8899384
## 90+                            0.9710000
## all 12+                        0.7460000
```

Part 3: Here, the age standardized vaccine efficacy is calculated

```
# Standardize the vaccine efficacy
efficacy_age1 = (1 - (phat_vax_1)/(phat_unvax_1)) * ((971478 + 2573903)/
(1302910+5634632)) #12-39
efficacy_age2 = (1 - (phat_vax_1/phat_unvax_2)) * ((145355 + 927214)/
```

```r
(1302910+5634632))  #40-49
efficacy_age3 = (1 - (phat_vax_3/phat_unvax_3)) * ((84545 +
747949)/(1302910+5634632))  #50-59
efficacy_age4 = (1 - (phat_vax_4/phat_unvax_4)) * ((65205 + 665717) /
(1302910+5634632)) #60-69
efficacy_age5 = (1 - (phat_vax_5/phat_unvax_5)) * ((20512 +
464336)/(1302910+5634632))  #70-79
efficacy_age6 = (1 - (phat_vax_6/phat_unvax_6)) * ((12683 +
208911)/(1302910+5634632))  #80-89
efficacy_age7 = (1 - (phat_vax_7/phat_unvax_7)) * ((3132 +
46602)/(1302910+5634632))   #90+
total = efficacy_age1 + efficacy_age2 + efficacy_age3 + efficacy_age4 +
      efficacy_age5 + efficacy_age6 + efficacy_age7
total
```

```
## [1] 0.9443367
```

```r
# Compare to the actual
efficacy_observed = 1 - (301/5634632)/(214/1302910)
efficacy_observed
```

```
## [1] 0.6747618
```

```r
diff = total - efficacy_observed
diff
```

```
## [1] 0.2695749
```