Placebo-Controlled Efficacy Analysis of the COVID-19 Vaccine

Spring 2024

Oliver Brown, Josie Czeskleba, Luke VanHouten

Abstract

Keywords

Efficacy, Inference, COVID-19, Statistics, Estimators

Introduction

In this project, we will be analyzing the efficacy of the Pfizer-BioNTech BNT162b2 COVID-19 mRNA vaccine based on a sample of placebo-controlled COVID tests (Polack et al. 2020). This data was collected in late 2020 among individuals at least sixteen years old who received two doses of the vaccine three weeks apart, with them being tested for COVID afterwards. The efficacy of the vaccine is extremely important, because the goal of vaccination efforts are to save lives. So, we perform statistical testing to identify whether or not the efficacy rate shown in the vaccine data is acceptable. Our hypothesis is that the BNT162B2 vaccine efficacy is 95%, which we will test using different statistical methods.

Statistical Methods

We denote the random variable T as the number of vaccinated individuals from the 170 COVID cases.

$$T \sim Binom(n = 170, \pi)$$

We can define $\pi = \text{P(Vaccine|COVID)} = \frac{\pi_1}{\pi_1+\pi_2}$, given that the sample sizes for the vaccine and placebo groups are approximately equal. Here, π_1 is the proportion of vaccinated individuals who got COVID and π_2 is the proportion of unvaccinated individuals who got COVID. Moreover, we define the vaccine efficacy as $\psi = \frac{1-2\pi}{1-\pi}$.

Maximum Likelihood Estimator

We can first write the likelihood function of π

$$L(\pi) = \binom{n}{t}\pi^t(1-\pi)^{n-t}$$

Then we write π in the form $\pi = g(\psi)$, given that $\psi = \frac{1-2\pi}{1-\pi}$. We thus have that $\psi - \psi\pi = 1 - 2\pi$, which becomes $2\pi - \psi\pi = 1 - \psi$, which becomes:

$$\pi = \frac{1-\psi}{2-\psi}$$

We can then write the likelihood function for ψ:

$$L(\psi) = L(g(\psi)) = L\left(\frac{1-\psi}{2-\psi}\right) = \binom{n}{t}\left(\frac{1-\psi}{2-\psi}\right)^t\left(1-\left(\frac{1-\psi}{2-\psi}\right)\right)^{n-t} = \binom{n}{t}\left(\frac{1-\psi}{2-\psi}\right)^t\left(\frac{1}{2-\psi}\right)^{n-t}$$

We can then calculate the log-likelihood function for ψ:

$$\ell(\psi) = \ln(L(\psi) = \ln\left(\binom{n}{t}\right) + t\ln(1-\psi) - t\ln(2-\psi) - (n-t)\ln(2-\psi) = \ln\left(\binom{n}{t}\right) + t\ln(1-\psi) - n\ln(2-\psi)$$

We can then find our estimator by setting $\ell'(\psi) = 0$:

$$\frac{d}{d\psi}\ell(\psi) = \frac{d}{d\psi}\ln\left(\binom{n}{t}\right) + \frac{d}{d\psi}t\ln(1-\psi) - \frac{d}{d\psi}n\ln(2-\psi) = \frac{n}{2-\psi} - \frac{t}{1-\psi} = 0$$

We can then solve $\frac{n}{2-\psi} = \frac{t}{1-\psi}$. We get that $n - n\psi = 2t - t\psi$, which becomes $t\psi - n\psi = 2t - n$, giving us an estimator of $\widehat{\psi}_0^{mle} = \frac{2t-n}{t-n}$.

Bootstrap

Results

For our MLE, we can plug in $t_{obs} = 8$ and $n = 170$ to $\widehat{\psi}_0^{mle} = \frac{2t-n}{t-n}$, we get $\widehat{\psi}_0^{mle} = \frac{16-170}{8-170} = \frac{77}{81} = 0.9506$. We can also use the Newton Raphson method to estimate ψ to get the same value, shown in the appendix.

Conclusion

References

Polack, Fernando P., Stephen J. Thomas, Nicholas Kitchin, Judith Absalon, Alejandra Gurtman, Stephen Lockhart, John L. Perez, et al. 2020. "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine." *New England Journal of Medicine* 383 (27): 2603–15. https://doi.org/10.1056/NEJMoa2034577.

Senn, Stephen. 2021. "S. Senn: 'Beta Testing': The Pfizer/BioNTech Statistical Analysis of Their Covid-19 Vaccine Trial (Guest Post)." *Error Statistics Philosophy*. https://errorstatistics.com/2021/01/17/s-senn-beta-testing-the-pfizer-biontech-statistical-analysis-of-their-covid-19-vaccine-trial-guest-post/.

Appendix

Newton Rhapson MLE Approximation

```r
loglik = function(psi, T, n){
    return(log(choose(n, T)) + (T * log(1 - psi)) - (n * log(2 - psi)))
}

maxLik(logLik = loglik, start = 0.55, method = "NR", tol = 1e-4, T = 8, n = 170)
```

```
## Maximum Likelihood estimation
## Newton-Raphson maximisation, 6 iterations
## Return code 2: successive function values within tolerance limit (tol)
## Log-Likelihood: -1.944994 (1 free parameter(s))
## Estimate(s): 0.9506174
```

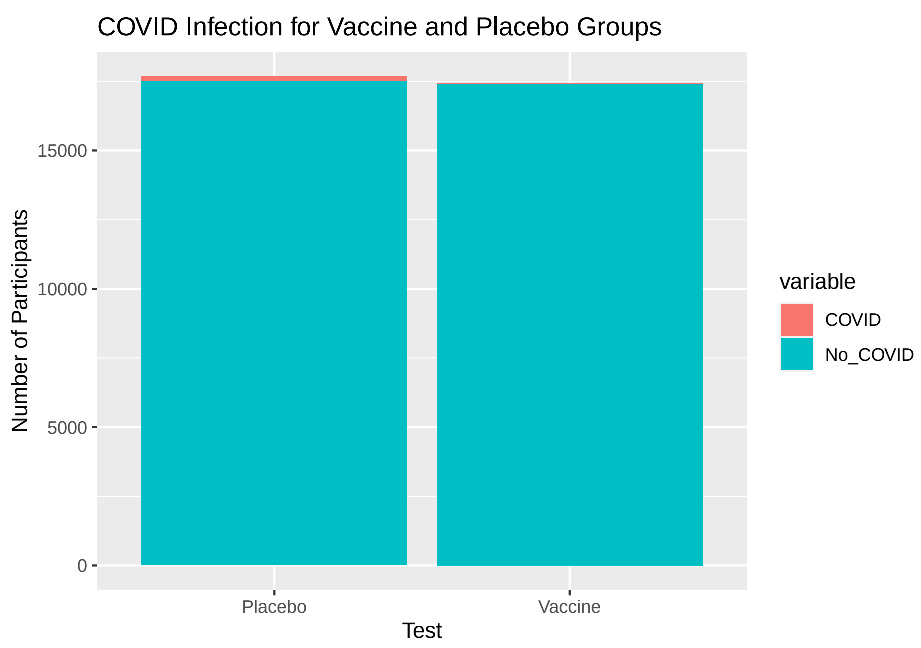
Visualizations

```r
data <- read.csv("data.csv")

data_melted <- melt(data, id.vars = "Test")
```

Stacked Barplot

```r
ggplot(data_melted, aes(x = Test, y = value, fill = variable)) +
  geom_bar(stat = "identity") +
  labs(x = "Test", y = "Number of Participants",
       title = "COVID Infection for Vaccine and Placebo Groups")
```



Faceted Barplot

```
ggplot(data_melted, aes(x = Test, y = value)) +
  geom_bar(stat = "identity") +
  facet_wrap(~ variable, scales = "free_y") +
  labs(x = "Test", y = "Number of Participants",
       title = "COVID Infection for Vaccine and Placebo Groups")
```