

A Critical Approach to Evaluating the Efficacy of the J&J BNT162b2 Vaccine (COVID)

Arnav, Maxx, Mohit

03-18-25

Abstract

Write your abstract here.

Keywords

Keyword 1, Keyword 2, Keyword 3, Keyword 4

Introduction / Background

Provide an introduction with background information.

Statistical Methods

Model

T is the random variable that defines the number of people who had the vaccine out of the people who tested positive of a total $n = 170$ cases.

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

$$\pi = P(\text{Vaccinated} | \text{Positive Covid Test})$$

$$\pi = \frac{n_1 \pi_v}{n_1 \pi_v + n_2 \pi_p}, \text{ Where } n_1 \approx n_2, \text{ the randomization is } 1 : 1$$

$$\pi_v = P(\text{Testing positive} | \text{Having vaccine})$$

$$\pi_p = P(\text{Testing positive} | \text{Having placebo})$$

$$\pi \frac{\pi_v}{\pi_v + \pi_p}$$

$$\psi = 1 - \frac{\pi_v}{\pi_p}$$

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

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

Our parameter of interest is $\psi = \frac{1-2\pi}{1-\pi}$. We are interested in the interval estimate of ψ , our $H_0 : \psi = 0.3$ and our $H_1 : \psi \neq 0.3$. We took two different approaches for analysis, which was Likelihood and Bayesian.

Likelihood Inference

Maximum likelihood estimators (MLE) are invariant to transformation so we find the MLE for π_0^{MLE} and then easily transform it to ψ_0^{MLE}

$$T \sim \text{Binom}(170, \pi_0) \leftrightarrow T_1, T_2, \dots, T_{170} \sim \text{Bernouli}(\pi_0)$$

$$\hat{\pi}_0^{MLE} = \frac{T}{n}$$

$$\hat{\psi}_0^{MLE} = \frac{1 - 2\pi_0}{1 - \pi_0} = \frac{n - 2T}{n - T}$$

Standard errors and significance tests are not invariant to transformations, which means the likelihood function will be written in terms of ψ

$$g(\psi) = \frac{1 - \psi}{2 - \psi}$$

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

$$\ell^*(\psi) = \log(L^*(\psi)) = \log\binom{n}{t} + t \cdot \log\left(\frac{1 - \psi}{2 - \psi}\right) + (n - t) \log\left(1 - \frac{1 - \psi}{2 - \psi}\right)$$

$$\frac{d}{d\psi} \ell^*(\psi) = \frac{n - t}{2 - \psi} - \frac{t}{(1 - \psi)(2 - \psi)}$$

$$0 = \frac{n - t}{2 - \psi} - \frac{t}{(1 - \psi)(2 - \psi)}, \text{ For MLE condition}$$

From there, we would need to do the second derivative test to make sure our value is a maximum:

$$\frac{d^2}{d\psi^2} \ell^*(\psi) = \frac{n - t}{(2 - \psi)^2} + \frac{t(2\psi - 3)}{((1 - \psi)(2 - \psi))^2}$$

A large sample MLE is approximately normally distributed:

$$\begin{aligned}
\hat{\psi}_0^{MLE} &\approx Norm(\psi_0, \sqrt{\frac{1}{nI(\psi_0)}}) \\
I(\psi_0) &= E\left[\frac{-d^2}{d\psi_0^2} \log(f_{\psi_0}(T))\right] \\
f_{\psi_0}(T) &= \binom{n}{t} \pi^t (1 - \pi)^{n-t} \\
f_{\psi_0}(T) &= \binom{n}{t} \left(\frac{1 - \psi}{2 - \psi}\right)^t \left(1 - \frac{1 - \psi}{2 - \psi}\right)^{n-t} = L^*(\psi) \\
\log(f_{\psi_0}(T)) &= \ell^*(\psi) \\
I(\psi_0) &= E\left[-\left(\frac{n - t}{(2 - \psi)^2} + \frac{t(2\psi - 3)}{((1 - \psi)(2 - \psi))^2}\right)\right] \\
CI_{\alpha=0.05}(\hat{\psi}_0^{MLE}) &= \hat{\psi}_0^{MLE} \pm Z_{\alpha/2} \cdot \sqrt{\frac{1}{nI(\psi_0)}}
\end{aligned}$$

$$H_0 : \psi_0 = 0.3$$

$$H_1 : \psi_0 \neq 0.3$$

$$w = 2\log(\Lambda)$$

$$\Lambda = \frac{L(\hat{\psi}_0^{MLE})}{L(\psi_{null})}$$

Bayesian Inference

For the bayesian inference, we first started by defining our prior probability of efficacy (ψ), which we chose overly pessimistic probabilities for our prior which means if we still get strong results, our evidence will be very compelling. Then we translate the prior probability to be in terms of π , then we find the posterior and retranslate it back to ψ . This allows us to find the confidence interval for π and ψ as well as the posterior probabilities for ψ , which would give us evidence on the vaccine efficacy rate and whether it meets the FDA requirements for efficacy.

Results

Likelihood

When we solved our equation for the MLE condition, our value for $\hat{\psi}_0^{MLE}$ was 0.9506. When we plugged it into our second derivative test, we got a value of -3123.83, which means the function is concave downwards at that point, so our value is a global maximum. For finding the standard error, we do:

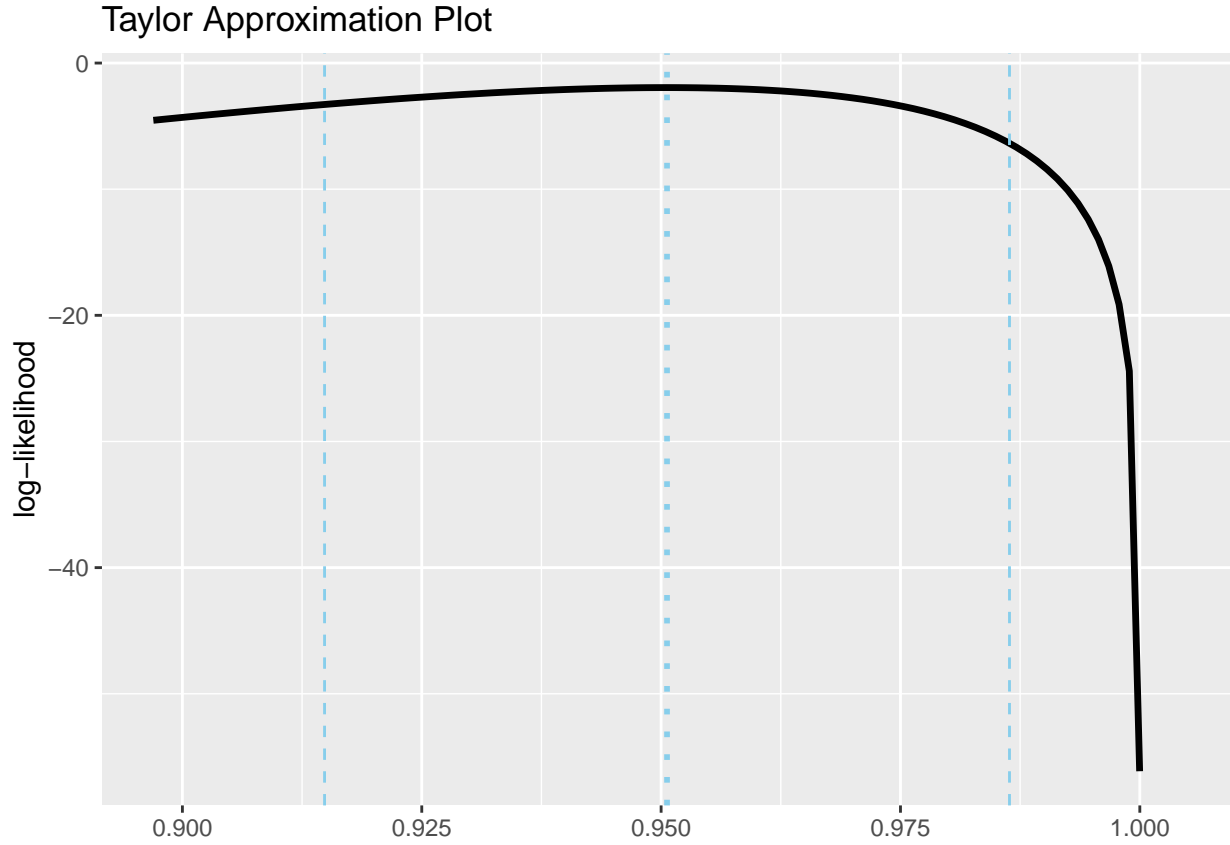
$$I(\psi_0) = E\left[\frac{-d^2}{d\psi_0^2}\log(f_{\psi_0}(T))\right]$$

$$I(0.9506) = 3123.8335$$

$$\sqrt{\frac{1}{nI(\psi_0)}} = 0.00137$$

$$CI_{\alpha=0.05}(\hat{\psi}_0^{MLE}) = 0.9506 \pm 1.96 \cdot 0.00137$$

$$CI_{\alpha=0.05}(\hat{\psi}_0^{MLE}) = [0.9479, 0.9533]$$



We also conducted a bootstrap estimate using random binomial samples for π and converting to it ψ . Using this technique, we found a 95% confidence interval of [0.9102564, 0.9820359].

Our w statistic we calculated was 121.72, which has a p-value of 2.658272×10^{-28} .

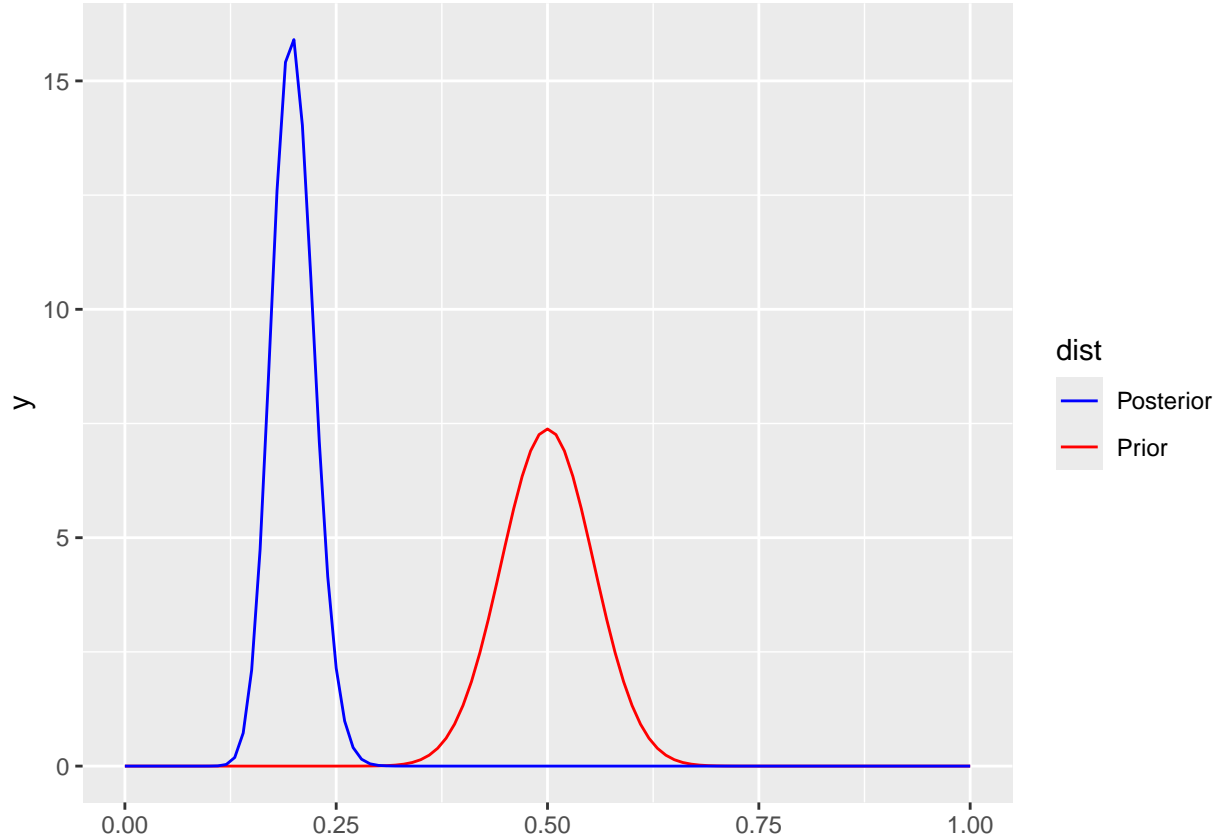
Bayesian

For the bayesian approach, we decided on a prior distribution of $Beta(43.03, 43.03)$. Which was decided on the probabilities:

$$P(\psi > 0) = 0.5$$

$$P(\psi > 0.3) = 0.05$$

From this we calculated our posterior distribution to be: $Beta(51.03, 205.03)$. These distributions are visualized below.



Using our posterior distribution for π , we found the median and confidence interval for π and translated it back to ψ . For the confidence interval we used the 95 percent highest posterior density interval. Our ψ median was 0.7523308, and the confidence interval was [0.6691396, 0.8217393]. And our p-value for our ψ median was $3.2446641 \times 10^{-77}$.

Discussion / Conclusion

Using likelihood inference, we are 95 percent confident the true value of ψ_0 (Vaccine Efficacy) is in the interval: [0.9479, 0.9533].

Using bootstrap, we are 95 percent confident that the true value for ψ_0 (Vaccine Efficacy) is in the interval: [0.9103, 0.9820].

Using Bayesian inference, we are 95 percent confident that the true value for ψ_0 (Vaccine Efficacy) is in the interval: [0.6691, 0.8217].

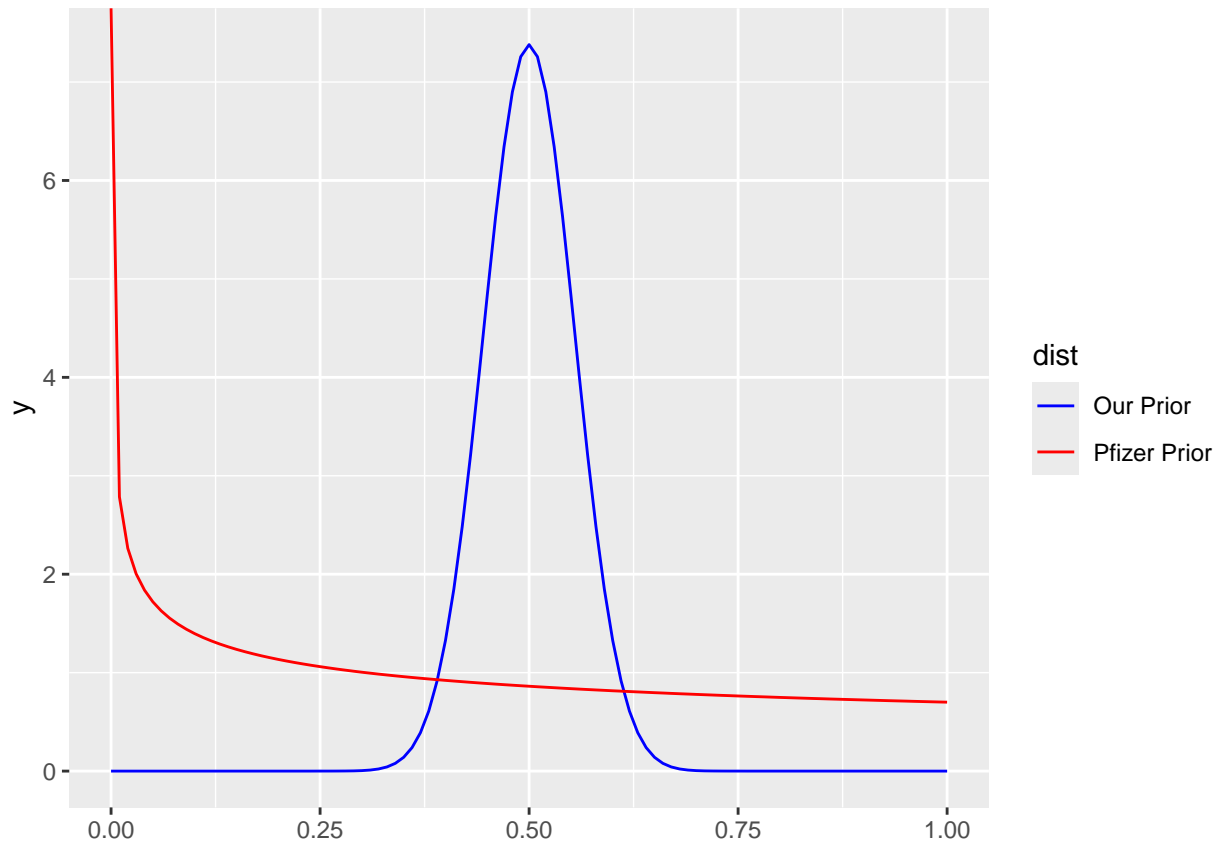
Our p-value for ψ using likelihood inference is: $2.6783 \cdot 10^{-28}$, which is significantly lower than our

significance value of $\alpha = 0.05$, so as a result we reject the null as there is convincing evidence that the true value for $\psi \neq 0.3$, which means that the efficacy of the vaccine is greater than the FDA requirement of 30 percent.

Our two-sided p-value for ψ using Bayesian inference is: $3.2447 \cdot 10^{-77}$, which is less than 0.05, therefore we have convincing evidence that the true value for the efficacy of the vaccine is different than the 30 percent FDA requirement.

Although we get different p-values and confidence intervals for the different techniques, ultimately they all show we have enough evidence that the true efficacy of the vaccine, that being the vaccine makes you less likely to contract COVID, is greater than the FDA requirement of 30 percent. The likelihood inference was a lot more optimistic as our confidence interval was a lot higher, whereas the Bayesian inference was a lot more pessimistic as our confidence interval was lower. This is because our prior distribution in the Bayesian inference was very critical only giving a 5 percent chance of exceeding FDA guidelines.

In comparison to the Pfizer study, their prior distribution of π for Bayesian inference was $Beta(0.700102, 1)$ compared to our $Beta(43.03, 43.03)$:



Which as we can see gives π a really high probability of being smaller, which increases the efficacy of the vaccine.

In general, a strength of the frequentist approach is that it does not rely on a prior probability/distribution which means it solely relies on the data. As a result, everyone that uses the same data will get the same result. In contrast, the Bayesian approach allows for the prior probability/distribution, which can be beneficial if you have prior knowledge. It also allows for a more critical approach if data is very strong one direction. The downside to this, is humans can bias the prior distribution to better reach a conclusion they are looking for.

Bibliography

Brown, B. (2024). *Lecture Title*. Lecture slides, Course Name, University Name.

Doe, J. (2020). Title of the Paper. *Journal Name*, 12(3), 45-67.

Last, F., & Last, F. (2025). *Book Title*. Publisher.

Smith, A., & Johnson, C. (2023). *Title of the Online Article*. Retrieved from <https://www.example.com>.

Appendix

Code

For Taylor Plot

```
library(fastR2)
loglikelihood <- function(psi){
  n <- 170
  t <- 8
  return(log(choose(n, t)) + t * log((1 - psi) / (2 - psi)) + (n - t) * log(1 - (1 - psi) / (2 - psi)))
}

ml.psi <- maxLik2(loglik = loglikelihood, start=0.3)
plot(ml.psi) + labs(title = "Taylor Approximation Plot")
```

For Bootstrap

```
set.seed(123)

n <- 170
psi_mle <- 0.9506
pi_hat <- (1 - psi_mle) / (2 - psi_mle)

n_bootstrap <- 10000

T_bootstrap <- rbinom(n_bootstrap, size = n, prob = pi_hat)

psi_bootstrap <- (n - 2 * T_bootstrap) / (n - T_bootstrap)
```

```

ci_boot <- quantile(psi_bootstrap, probs = c(0.025, 0.975))

boot_lower <- unname(ci_boot[1])
boot_upper <- unname(ci_boot[2])

```

For Prior-Posterior Plot

```

a <- 8 + 43.03
b <- 170 - 8 + 43.03

library(tidyverse)

ggplot() +
  geom_function(fun = dbeta,
               mapping = aes(color = 'Prior'),
               args = list(shape1 = 43.03, shape = 43.03),
               xlim = c(0, 1)) +
  geom_function(fun = dbeta,
               mapping = aes(color = 'Posterior'),
               args = list(shape1 = a, shape2 = b),
               xlim = c(0, 1)) +
  scale_color_manual(name = 'dist', values = c('blue', 'red'))

```

For Bayesian Calculations

```

pi_median <- qbeta(0.5, a, b)
pi_ci <- qbeta(p = c(0.025, 0.975), a, b)

psi_median <- (1 - 2 * pi_median) / (1 - pi_median)

```

```

library(HDInterval)

hd_pi_ci <- hdi(qbeta, credMass=0.95, shape1 = a, shape2 = b)
pi_lower <- unname(hd_pi_ci[1])
pi_upper <- unname(hd_pi_ci[2])
psi_upper <- (1 - 2 * pi_lower) / (1 - pi_lower)
psi_lower <- (1 - 2 * pi_upper) / (1 - pi_upper)
psi_median_p_value <- 2 * pbeta(psi_median, a, b, lower.tail = FALSE)

```

For Prior Comparison

```

ggplot() +
  geom_function(fun = dbeta,
               mapping = aes(color = 'Our Prior'),
               args = list(shape1 = 43.03, shape2 = 43.03),
               xlim = c(0, 1)) +
  geom_function(fun = dbeta,
               mapping = aes(color = 'Pfizer Prior'),
               args = list(shape1 = 0.700102, shape2 = 1),
               xlim = c(0, 1)) +
  scale_color_manual(name = 'dist', values = c('blue', 'red'))

```

Proofs

If applicable, include detailed mathematical derivations or additional theoretical explanations.