Analyzing COVID-19 Vaccine Efficacy: Comparing MLE and MOM Methods for Pfizer-BioNTech's BNT162b2

Alexandra Orczyk, Vy Nguyen, Xiyue Huang

1 Abstract:

Background: Prior studies found BNT162b2 vaccine efficacy to be 95%, with a Bayesian interval of [90.3%, 97.6%], and a probability exceeding 30% to be over 0.9999 (Polack, et. al). Building on these findings, we will conduct an analysis using different methods to test the efficacy of the vaccine.

Method: To evaluate the efficacy of the vaccine, we utilize likelihood function and method of moments with a binomial distribution to establish hypothesis testing and calculate confidence intervals by bootstrapping

Results: In both cases, the p-value approached 0, indicating sufficient evidence to reject the null hypothesis that efficacy rate is at 30%. Efficacy rates were calculated to be between 91.48% and 98.65%, indicating strong evidence supporting the reduction in symptomatic COVID-19.

Conclusion: The majority of studies confirm the high efficacy of BNT162b2 vaccine, demonstrating their safety and effectiveness in preventing infection and reducing symptom severity.

2 Keywords:

COVID-19, BNT162b2, Vaccine efficacy, Binomial distribution, Hypothesis testing, Confidence Interval

3 Introduction:

Due to the critical importance of reducing infections during the COVID-19 pandemic, and the critical role that vaccines could play in this task, it is extremely important to determine how effective different vaccines are in preventing COVID. Prior studies on the efficacy of the various COVID-19 vaccines, particiularly the Pfizer and Moderna mRNA vaccines, have found them to be highly effective at lowering infection, lowering the severity of symptoms, and preventing hospitalization and death. For example, Cai, Peng, et.al calculated a 95% confidence interval of 93.65%–95.40% as an estimate of the vaccine efficacy rate of the mRNA vaccines, which was defined as $\frac{\text{Risk among unvaccinated group} - \text{risk among vaccinated group}}{\text{Risk among unvaccinated group}}$. Similarly, a systematic review by Grana, Ghosn, et. al found "high-certainty evidence" that the BioNTech/Pfizer vaccine reduced both the prevalence of symptomatic COVID-19 and symptom severity compared to placebo. However, they found lower efficacy against newer COVID-19 variants. Although there are some estimates that make the vaccines seem less effective, such as the Absolute Risk Reduction of 0.9% for the Pfizer vaccine reported by Olliaro, Torreele, et.al, however, since this statistic reports the difference in infection rates between the unvaccinated and vaccinated population, it has the flaw of reporting a higher degree of efficacy for the vaccines tested in populations with a higher rate of Covid-19, which is why most studies use other estimators. Overall, the majority of studies report a high degree of efficacy for the mRNA Covid-19 vaccines. In this study, our goal is to examine the efficacy of the Pfizer-BioNTech Covid-19 vaccines through two different statistical methods: the maximum likelihood estimate and the method of moments. Because of the high prior consensus on the efficacy of the vaccines, we believe that both methods will show a significant effect of the vaccine. However, in line with the previous study done by Polack et. al., we will use a null hypothesis that the vaccine efficacy is equal to 30%. Using both the maximum likelihood and the method of moments, we will conduct hypothesis testing and create confidence intervals to estimate the efficacy of the vaccines.

4 Statistical Methods:

Let X denote the number of COVID-19 cases in the BNT162b2 (vaccine) group out of the total n=170 COVID-19 cases observed in the study. We assume that X follows a binomial distribution:

$X \sim Binom(n = 170, \pi)$, where π is the probability of observing a COVID-19 case in the vaccine group. In this case, $\pi = \frac{\pi_v}{\pi_v + \pi_p}$ where π_v and π_p are the probabilities of an infection in the vaccine and placebo groups respectively.

The parameter of interest is the vaccine efficacy, denoted by ψ. Vaccine efficacy is related to the binomial probability π through the following relationship:

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

We can also express π as a function of:

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

4.1 Method 1: Likelihood Inference for Vaccine efficacy

To estimate the vaccine efficacy ψ, we will use the likelihood function for π and then transform it to infer ψ. The likelihood function for π given the observed data x is: $L(\pi) = \binom{n}{x}\pi^x.(1 - \pi)^{n-x}$.

Since our parameter of interest is ψ, we substitute $\pi = \frac{1-\psi}{2-\psi}$ into the likelihood function to get: $L(\psi) = L(g(\psi)) = \binom{n}{x}(\frac{1-\psi}{2-\psi})^x.(1 - \frac{1-\psi}{2-\psi})^{n-x}$

We can simplify the expression for the likelihood function $L(\psi)$:

$$\begin{aligned}
L(\psi) &= \binom{n}{x}(\frac{1-\psi}{2-\psi})^x.(1 - \frac{1-\psi}{2-\psi})^{n-x} \\
&= \binom{n}{x}(\frac{1-\psi}{2-\psi})^x.(\frac{2-\psi-1+\psi}{2-\psi})^{n-x} \\
&= \binom{n}{x}(1-\psi)^x(\frac{1}{2-\psi})^x.(\frac{1}{2-\psi})^{n-x} \\
&= \binom{n}{x}\frac{(1-\psi)^x}{(2-\psi)^n} \qquad 0 \leq \psi \leq 1
\end{aligned}$$

The log-likelihood function is then given by: $l(\psi) = ln\binom{n}{x} + xln(1 - \psi) - nln(2 - \psi)$

To find ψ^{mle}, we will derive the first derivative of the log-likelihood function:

$$\begin{aligned}\frac{d}{d_\psi}l(\psi) &= \frac{d}{d_\psi}(ln\binom{n}{x} + xln(1-\psi) - nln(2-\psi)) \\ &= \frac{-x}{1-\psi} + \frac{n}{2-\psi} \qquad 0 \leq \psi < 1\end{aligned}$$

Candidates for the MLE of ψ satisfy the equation:

$$\begin{aligned}\frac{d}{d_\psi}l(\psi_0) &= 0 \\ \frac{-x}{1-\psi_0} + \frac{n}{2-\psi_0} &= 0 \\ \frac{n}{2-\psi_0} &= \frac{x}{1-\psi_0} \\ n(1-\psi_0) &= x(2-\psi_0) \\ n - n\psi_0 &= 2x - x\psi_0 \\ n - 2x &= (n-x)\psi_0 \\ \psi_0^{mle} &= \frac{n-2x}{n-x}\end{aligned}$$

We need to use second derivative test of the log-likelihood function to ensure concavity:

$$\begin{aligned}\frac{d^2}{d_\psi^2}l(\psi_0) &= \frac{d}{d_\psi}(\frac{-x}{1-\psi_0} + \frac{n}{2-\psi_0}) \\ &= \frac{x}{(1-\psi_0)^2} - \frac{n}{(2-\psi_0)^2} \\ &= \frac{-x}{(1-\frac{n-2x}{n-x})^2} + \frac{n}{(2-\frac{n-2x}{n-x})^2} \\ &= \frac{-x}{(\frac{n-x-n+2x}{n-x})^2} + \frac{n}{(\frac{2n-2x-n+2x}{n-x})^2} \\ &= \frac{-x}{(\frac{x}{n-x})^2} + \frac{n}{(\frac{n}{n-x})^2} \\ &= \frac{-1}{\frac{x}{(n-x)^2}} + \frac{1}{\frac{n}{(n-x)^2}} \\ &= (n-x)^2(\frac{1}{n} - \frac{1}{x})\end{aligned}$$

As we know that $x \leq n$, then $(n-x)^2 \geq 0$ and $\frac{1}{n} \leq \frac{1}{x} = \frac{1}{n} - \frac{1}{x} \leq 0$. Thus, $l"(\psi_0^{mle}) =$

$(n-x)^2(\frac{1}{n}-\frac{1}{x}) \leq 0$, so ψ_0^{mle} is the maximum likelihood estimate of ψ_0.

Now, we want to have two different statistical inferences for the vaccine efficacy - confidence intervals and (empirical) P-value.

Confident Interval of ψ_0^{mle}

By theorem 25.1, we know for large values of n, ψ_0^{mle} is approximately normally distributed with mean ψ_0 and estimated standard error given by $\sqrt{\frac{1}{l"(\psi_0^{mle})}}$

Now we can derive the standard error of ψ_0

$$\begin{aligned} SE(\psi_0) &= \sqrt{\frac{-1}{l"(\psi^{mle})}} \\ &= \sqrt{\frac{(1-\psi_0)^2(2-\psi_0)^2}{x(2-\psi_0)^2 - n(1-\psi_0)^2}} \\ &= \frac{(1-\psi_0)(2-\psi_0)}{\sqrt{x(2-\psi_0)^2 - n(1-\psi_0)^2}} \end{aligned}$$

So, the confidence interval for ψ_0 is: $\hat{\psi_0}^{mle} \pm z_{\alpha/2}SE = \hat{\psi_0}^{mle} \pm z_{\alpha/2} \cdot \frac{(1-\psi_0)(2-\psi_0)}{\sqrt{-x(2-\psi_0)^2+n(1-\psi_0)^2}}$

Hypothesis Testing for ψ_0^{mle}

Let ψ_0 denote the true (but unknown) value of vaccine efficacy. To test the vaccine efficacy, we set up the hypotheses:

- Null Hypothesis $(H_0) : \psi_0 = \psi_0^{null}$,
- Alternative Hypothesis $(H_1) : \psi_0 \neq \psi_0^{null}$

We will use the likelihood ratio test to calculate the p-value since it is better than large sample z test when the sample size is moderate. The likelihood ratio test statistic follows a chi-square distribution with 1 degree of freedom:

$$W = 2ln[\frac{L(\hat{\psi_0}^{mle})}{L(\psi_0^{null})}] = 2ln[l(\psi_0^{mle}) - l(\psi_0^{null})] \approx X_1^2$$

If the p-value is less than the significance level α, we reject the null hypothesis and conclude that the BNT162b2 vaccine has a significant effect.

4.2 Method 2: Method of Moment Inference for Vaccine efficacy

We are interested in the value of $\hat{\psi}_0^{mom}$. Since $\pi = \frac{1-\psi}{2-\psi}$, we can find the M.O.M estimate of ψ_0 by solving the equation $E(X) = \bar{x}$, where $\bar{x}$ equals to x in our data.

$$
\begin{aligned}
E(X) &= x \\
n\pi &= x \\
n\frac{1-\psi}{2-\psi} &= x \\
2x - x\psi &= n - n\psi \\
\psi(n-x) &= n - 2x \\
\psi &= \frac{n-2x}{n-x}
\end{aligned}
$$

Therefore, $\hat{\psi}_0^{mom} = \frac{n-2x}{n-x}$. Now, in order to make inferences about ψ (the vaccine efficacy) using the method of moments (MOM) approach, we can construct confidence intervals (CI) and conduct hypothesis testing.

Confident Interval of ψ_0^{mom}

To construct confidence intervals for ψ, we will use the bootstrap method to construct confidence intervals for ψ: 1. Re-sampling: Randomly re-sample the original data with replacement. 2. Statistic Calculation: For each re-sample, calculate the MOM estimate ψ_0^{mom}. 3. Repetition: Repeat the re-sampling process 100 times to build a distribution of the MOM estimates. 4. Confidence Interval: Calculate confidence intervals of ψ: $\hat{\psi}_0^{mom} \pm Z_{\alpha/2} SE(\hat{\psi}_0^{mom})$ where α is the significant level.

Hypothesis Testing for ψ_0^{mom}

Another statistical inference of our estimation is hypothesis testing. In the context of vaccine efficacy, let ψ_0 denote the true (but unknown) value of vaccine efficacy. To test the vaccine efficacy, we set up the hypotheses:

• Null Hypothesis $(H_0) : \psi_0 = \psi_0^{null}$,

• Alternative Hypothesis $(H_1) : \psi_0 \neq \psi_0^{null}$

By using the empirical p-value from the bootstrap distribution, we can make a more robust inference about the vaccine efficacy ψ. If the empirical p-value is less than the significance level, we reject the null hypothesis and conclude that the vaccine has a significant effect.

5 Result

The clinical trial compared the number of COVID-19 cases in participants receiving the BNT162b2 (Pfizer) vaccine versus a placebo. Out of the total 170 COVID-19 cases observed in the study, 8 cases were in the BioNTech-Pfizer group, while 162 cases were in the placebo group. The proportions of subjects experiencing COVID-19 were significantly lower in the vaccinated group. Figure 1 displays the number of COVID-19 cases in each group, we can see BNT162b2 group had significantly fewer cases compared to the placebo group.

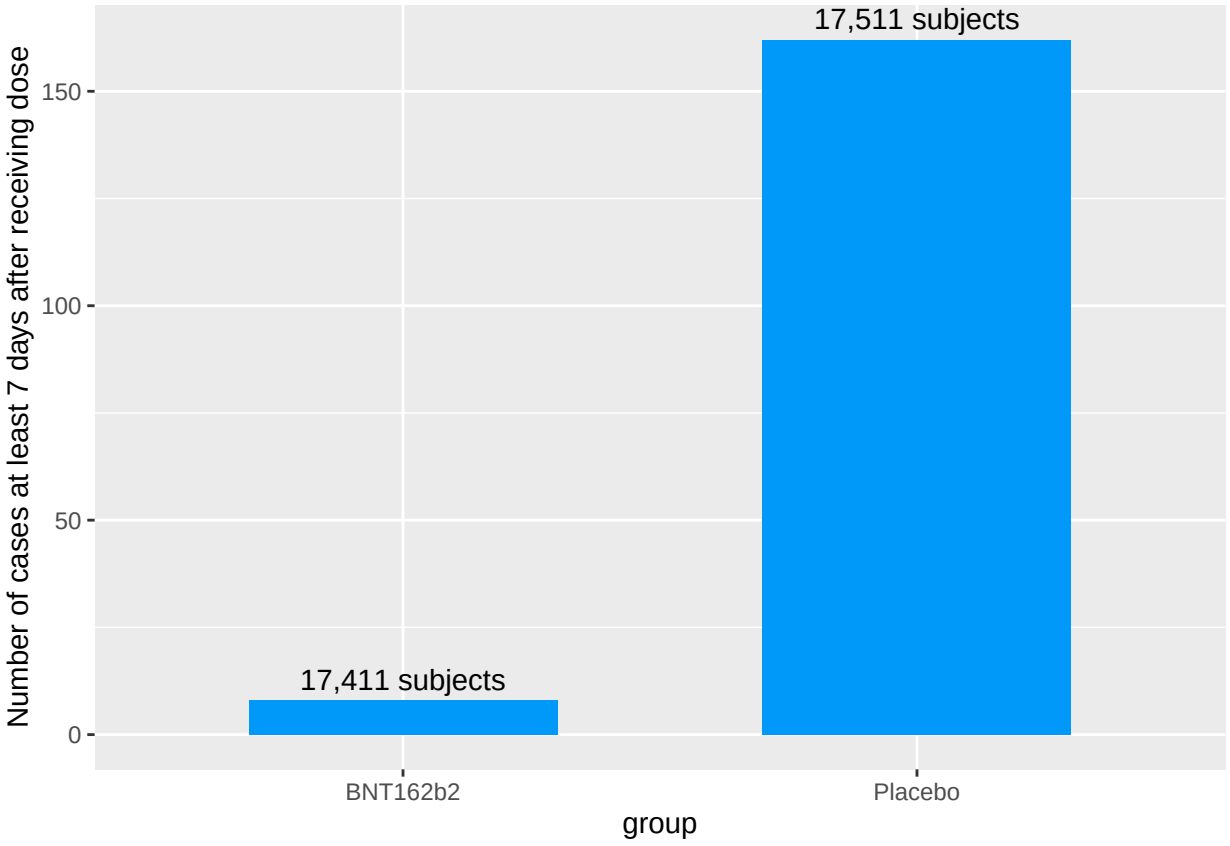

Figure 1. Number of COVID-19 cases in the group receiving the BNT162b2 vaccine versus the

placebo group.

5.1 Likelihood Inference for Vaccine Efficacy

Given the total number of infections $n = 170$ and the number of cases in the vaccine group $x = 8$, the maximum likelihood estimate (MLE) for vaccine efficacy ψ is

$$\psi_0^{mle} = 0.9506$$

The log-likelihood function for different values of ψ is shown in Figure 2, with the maximum likelihood estimate marked.

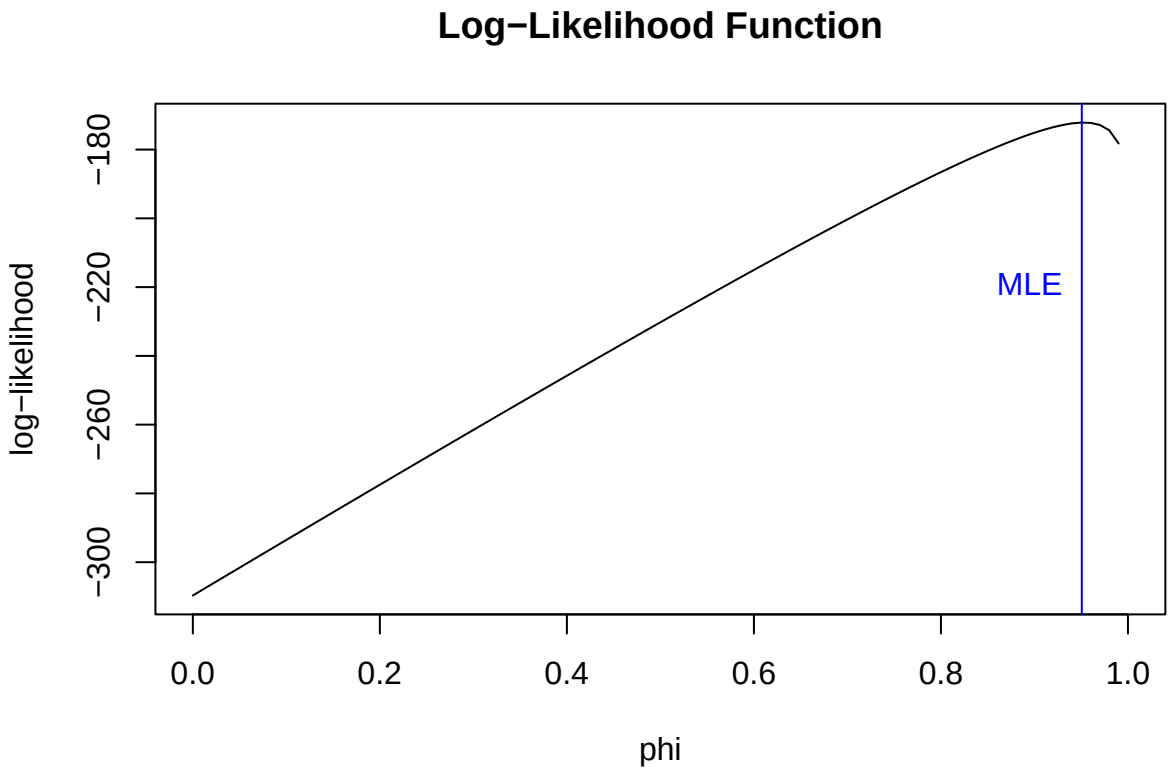

Figure 2. Log likelihood function and MLE

We used to theorem 25.1 to calculate standard error (SE) of the MLE. The theorem is hold under the regularity condition. Below is the graph to check the regularity condition:

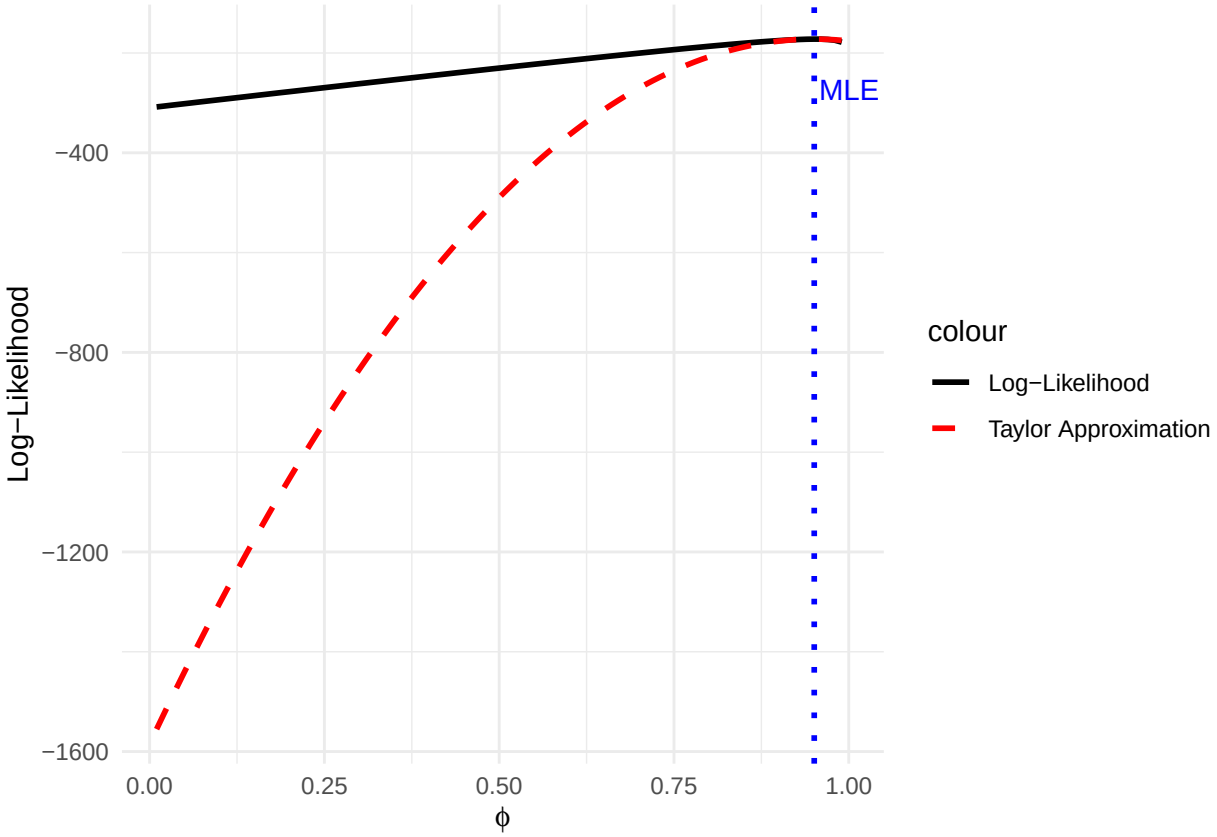

Figure 3. Log-Likelihood and Second-Order Taylor Approximation.

Therefore, we get the standard error (SE) of the MLE is:

$$SE(\psi) = 0.0178$$

The 95% confidence interval for ψ is:

$$[0.9156, 0.9857]$$

For the hypothesis testing, we are interested in the concern that if the probability that vaccine efficacy is 30%. So we set the test:

- Null Hypothesis $(H_0) : \psi_0 = 0.3$,
- Alternative Hypothesis $(H_1) : \psi_0 \neq 0.3$

Our test statistic is:

$$W = 2ln[\frac{L(\hat{\psi_0}^{mle})}{L(\psi_0^{null})}] = 121.6012$$

We get the P-value under the null hypothesis is 2.822294×10^{-28} and the empirical P-value by simulation is 0, confirming the significance with an empirical p-value is also less than 0.001. It can also be proved since the observed W falls in the tail of the null distribution, it suggests that the observed data is unlikely under the null hypothesis, providing evidence to reject the null hypothesis in figure 4.

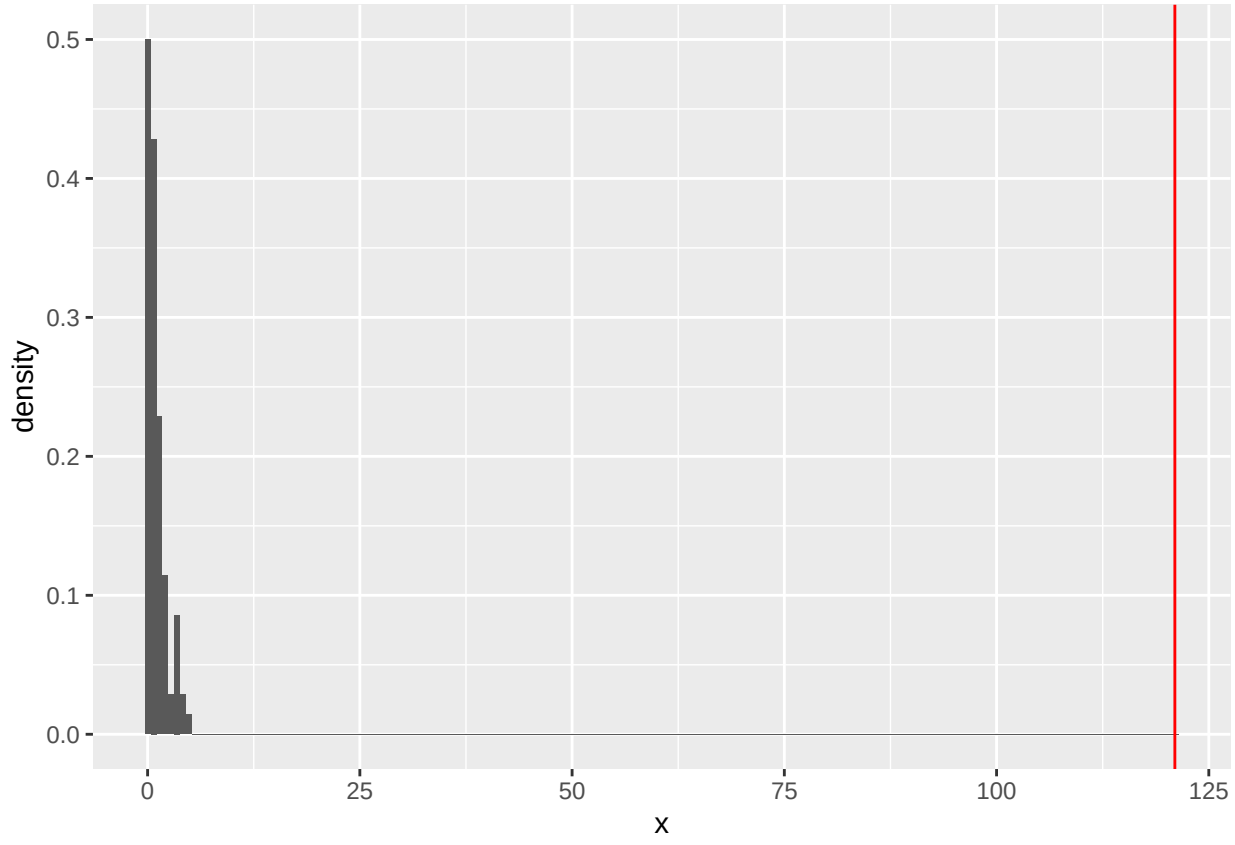

Figure 4. Null distribution of Test Statistic W.

5.2 Method of Moment Inference for Vaccine efficacy

Given the total number of infections $n = 170$ and the number of cases in the vaccine group $x = 8$, the MOM estimate for vaccine efficacy ψ is:

$$\psi_0^{mom} = \frac{n - 2x}{n - x} = \frac{170 - 16}{170 - 8} = 0.9506$$

Using bootstrapping, the 95% confidence interval for ψ_0^{mom} is:

$$[0.9148, 0.9865]$$

Figure 5 illustrates the 95% confidence intervals for the vaccine efficacy estimate over 100 bootstrap samples, with the observed Method of Moments estimate represented by a red dashed line. The red line falls within all of the bootstrap samples' blue intervals. This shows that the observed MOM estimate is relatively robust.

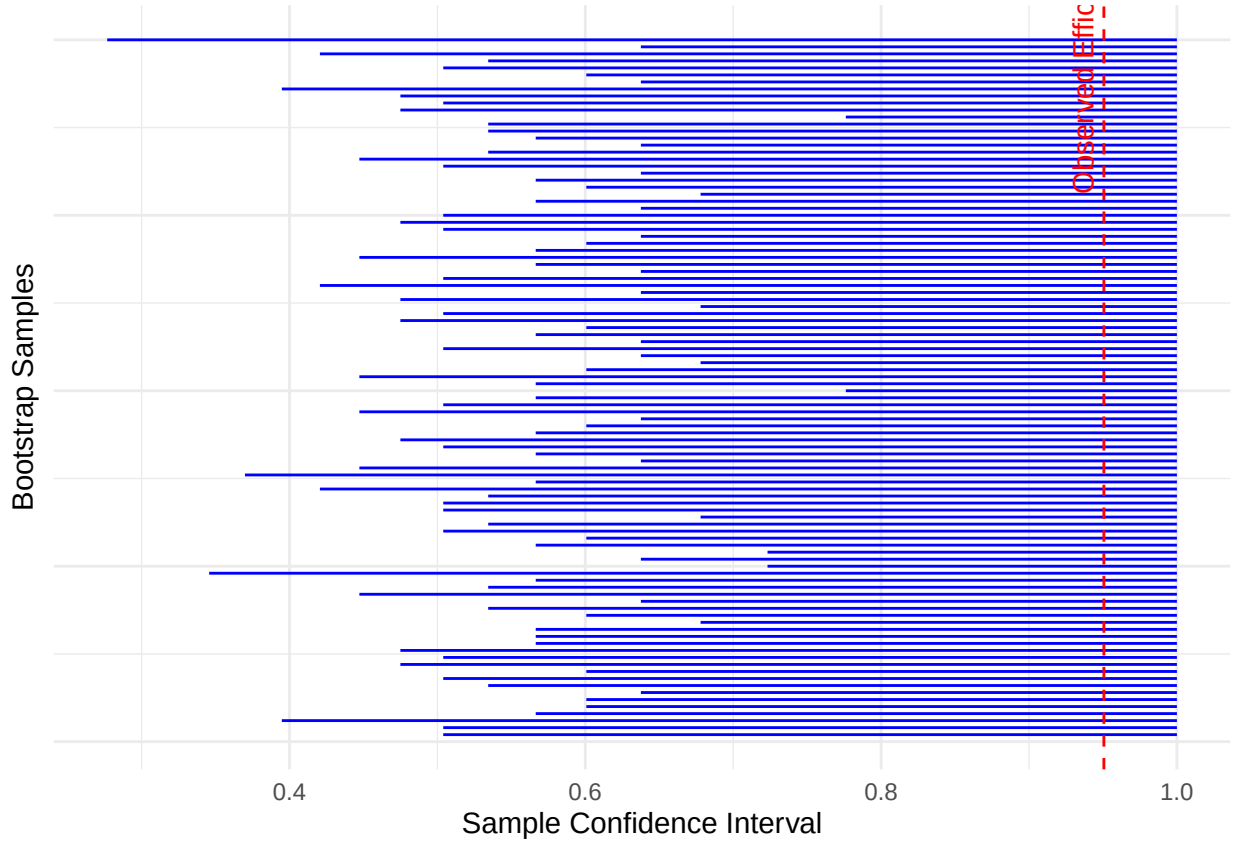

Figure 5. 95% Confidence Interval for 100 Bootstrap Samples.

For the hypothesis testing, we set the test:

• Null Hypothesis $(H_0) : \psi_0 = 0.3$,

• Alternative Hypothesis $(H_1) : \psi_0 \neq 0.3$

We get the empirical P-value by bootstrap equals to 0.

6 Discussion/Conclusion:

Our maximum likelihood estimate of the vaccine efficacy ψ was 0.9506. As the likelihood ratio test gave an extremely low p-value of 2.822294×10^{-28}, we can confidently reject the null hypothesis that $\psi = 0.3$. Under the method of empirical p-values, we calculated a p-value that is extremely close to 0, and in the plot we can see that our calculated value of W is very far away from any of the values of W under the null distribution.

However, one limitation of the maximum likelihood estimate is that the approximate chi square distribution for W relies on the second order approximation to the log likelihood being accurate. Since figure 3 shows that the second order approximation does not fit the log likelihood function very well, this indicates that the regularity condition is not fully met, which can impact the reliability of our confidence intervals and hypothesis tests.

Since the sample size is not very large, we also decided to calculate the P-value empirically through simulation, which is reliable even when there are smaller sample sizes.

Our method of moments estimate of ψ was also $\frac{n-2x}{n-x} = \frac{170-2\cdot 8}{170-8} = 0.9506$, with a 95% confidence interval of $[0.9148, 0.9865]$. When we bootstrapped 100 random samples, we found that none of of the confidence intervals contained $\psi_{null} = 0.3$, and we calculated a p-value that is extremely close to 0. As this p-value is far below our $\alpha = 0.05$, we have very evidence for rejecting the null hypothesis that the vaccine efficacy rate $\psi = 0.3$.

Comparing our results to the Pfizer analysis, although their method of using a Bayestian estimator was similar, we found similar confidence intervals for the vaccine efficacy. While the Pfizer study did not use p-values, they found a "more than a 99.99% probability of a true vaccine efficacy greater

than 30%". Conceptually, this is similar to the p-values for a vaccine efficacy $\neq 30\%$ we found that were extremely close to 0. So although we used different methodologies, our results strongly support the conclusion of the Pfizer study that was used in support of approval of the Pfizer-BionNTech vaccines.

In conclusion, as both the maximum likelihood estimate and method of moments show p-values very close to 0, we can conclude that the COVID-19 vaccines are very likely to have a high efficacy rate, most likely in the range of 91.48% to 98.65%. For practical policy implications, our results strongy suggest that public health systems should promote the distribution and use of the Pfizer-BionNTech mRNA vaccines as a tool against the Covid-19 Pandemic.

7 References:

Cai, Changjing, et al. "A comprehensive analysis of the efficacy and safety of COVID-19 vaccines." Cell, vol. 29, no. 9, 5 Aug. 2021, https://doi.org/10.1016/j.ymthe.2021.08.001.

Grana, Carolina, et al. "Efficacy and safety of COVID-19 vaccines." Cochrane Library, 7 Dec. 2022, https://doi.org/10.1002/14651858.CD015477.

Olliaro, Piero, et al. "COVID-19 vaccine efficacy and effectiveness—the elephant (not) in the room." The Lancet, vol. 2, no. 7, 20 Apr. 2021, https://doi.org/10.1016/S2666-5247(21)00069-0.

Polack, Fernando P., et al. "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine." The New England Journal of Medicine, vol. 383, no. 27, 10 Dec. 2020, doi:10.1056/NEJMoa2034577.

8 Appendix:

Regularity Conditions and Taylor Approximation

Theorem 25.1 can be used under certain regularity conditions. When n is large, and the regularity conditions hold, the log-likelihood function can be well approximated by a second order Taylor series around the MLE. Below is the R code to visualize this approximation:

```r
# Define the second-order Taylor approximation function
taylor_approx <- function(phi, x, n, phi_hat) {
  loglik_prime <- n / (2 - phi_hat) - x / (1 - phi_hat)
  loglik_double_prime <- n / (2 - phi_hat)^2 - x / (1 - phi_hat)^2
  return(loglik.binom(phi_hat, x, n) + loglik_prime * (phi - phi_hat)
         + 0.5 * loglik_double_prime * (phi - phi_hat)^2)
}

phi_values <- seq(0.01, 0.99, by = 0.01)
# Compute the log-likelihood values and the second-order Taylor approximation
loglik_values <- loglik.binom(phi_values, 8, 170)
taylor_values <- taylor_approx(phi_values, 8, 170, theta_0_mle_estimate)

# Create a dataframe for plotting
df <- data.frame(phi = phi_values, loglik = loglik_values, taylor = taylor_values)

# Plot
ggplot(df, aes(phi)) +
  geom_line(aes(y = loglik, color = "Log-Likelihood"), size = 1) +
  geom_line(aes(y = taylor, color = "Taylor Approximation"),
            linetype = "dashed", size = 1) +
  geom_vline(xintercept = theta_0_mle_estimate, color = "blue",
             linetype = "dotted", size = 1) +
  labs(title = "Log-Likelihood and Second-Order Taylor Approximation",
       x = expression(phi),
       y = "Log-Likelihood") +
  scale_color_manual(values = c("black", "red"),
                     labels = c("Log-Likelihood", "Taylor Approximation")) +
  theme_minimal() +
```

```
  annotate("text", x = theta_0_mle_estimate + 0.05,
           y = max(df$loglik, na.rm = TRUE) - 100,
           label = "MLE", color = "blue")
```

the distribution of test statistic W under the null hypothesis By comparing the observed W to the null distribution, we can assess whether the observed value is extreme or not. Below is the R code to visualize the null distribution of W:

```
# Define parameters
set.seed(123)
B <- 100
obs_w = 121
psi_null = 0.3
n = 170
x = 8
null_wstar = numeric(B)

### probability of COVID with vaccine
cases_vaccine = (1-psi_null)/(2-psi_null)

### probability of COVID with no vaccine
cases_placebo = 1-cases_vaccine

for(i in 1:B) {
  sample_boot <- sample(c(1, 0), size = n, replace = TRUE, prob = c(cases_vaccine, cases_placeb
  xstar <- sum(sample_boot == 1)
  mle <- (n - 2 * xstar) / (n - xstar)
  loglik_mle <- log(choose(n, xstar)) + xstar*log(1 - mle) - n*log(2 - mle)
  loglik_null <- log(choose(n, xstar)) + xstar*log(1 - psi_null) - n*log(2 - psi_null)
  null_wstar[i] <- 2 * (loglik_mle - loglik_null)
```

```
}


# Visualize the null distribution of wstar
ggplot(data = data.frame(x = null_wstar),
       mapping = aes(x = x)) +
  geom_histogram(mapping = aes(y = after_stat(density)),
                 binwidth = 0.7) +
  geom_vline(xintercept = obs_w, color = "red") +
  labs(title = "Null distribution of W")
```

P-value Calculation for MLE Method

We use the test statistic W to calculated the P-value.

```
# Define parameters
n = 170
x = 8
null = 0.3

# Calculate the mle value of psi
mle <- (170 - 2*8)/(170-8)

# Calculate the standard error of psi
se <- (1-mle)*(2-mle)/sqrt(8*(2-mle)^2 - 170*(1-mle)^2)

# Calculate confidence interval
z <- qnorm(0.975)
lower_mle <- mle - z*se
upper_mle <- mle + z*se

# The log likelihood function
```

```r
L_mle <- choose(n ,x)*(1-mle)^x * (1/(2-mle))^n
L_null <- choose(n ,x)*(1-null)^x * (1/(2-null))^n


# Calculate the test statistic W
w <- 2*log(L_mle/L_null)


# Calculate the p-value
pvalue = pchisq(w, df=1, lower.tail = F)
```

Bootstrap Sampling and Empirical P-value Calculation for MLE Method

Since our sample size is not very large, we calculated the empirical P-value by simulation.

```r
# Define parameters
set.seed(123)
obs_w = 121
B <- 100
psi_null = 0.3
n = 170
x = 8
null_wstar = numeric(B)


### probability of COVID with vaccine
cases_vaccine = (1-psi_null)/(2-psi_null)


### probability of COVID with no vaccine
cases_placebo = 1-cases_vaccine


for(i in 1:B) {
  sample_boot <- sample(c(1, 0), size = n,
                        replace = TRUE, prob = c(cases_vaccine, cases_placebo))
```

```
  xstar <- sum(sample_boot == 1)
  mle <- (n - 2 * xstar) / (n - xstar)
  loglik_mle <- log(choose(n, xstar)) + xstar*log(1 - mle) - n*log(2 - mle)
  loglik_null <- log(choose(n, xstar)) + xstar*log(1 - psi_null) - n*log(2 - psi_null)
  null_wstar[i] <- 2 * (loglik_mle - loglik_null)
}

# Visualize the null distribution of wstar
ggplot(data = data.frame(x = null_wstar),
       mapping = aes(x = x)) +
  geom_histogram(mapping = aes(y = after_stat(density)),
                 binwidth = 0.7) +
  geom_vline(xintercept = obs_w, color = "red") +
  labs(title = "Null distribution of W")

# Calculate empirical P-value by simulation
pvalue <- mean(null_wstar >= obs_w)
```

Bootstrap Sampling and Confidence Interval Calculation for MOM Method

We calculate the standard error across all bootstrap samples and compute the 95% confidence interval for the method of moments estimate.

```
# Define parameters
B <- 100
n <- 170
prob1 <- 8 / 170
prob2 <- 162 / 170
z <- qnorm(0.975)

# Initialize vectors to store bootstrap values and confidence intervals
```

```r
boot_values <- numeric(B)
se_boot <- numeric(B)
lower_mom_boot <- numeric(B)
upper_mom_boot <- numeric(B)

# Perform bootstrap sampling
set.seed(123)
for (i in 1:B) {
  sample_boot <- sample(c(1, 0), size = n,
                        replace = TRUE, prob = c(prob1, prob2))
  x <- length(which(sample_boot == 1))
  boot_values[i] <- (n - 2*x) / (n - x)
  se_boot[i] <- sd(sample_boot)  # Standard error for each bootstrap sample
  lower_mom_boot[i] <- boot_values[i] - z * se_boot[i]
  upper_mom_boot[i] <- pmin(boot_values[i] + z * se_boot[i],1)
}

# Calculate overall standard error for the estimator
se <- sd(boot_values)

# Calculate overall confidence interval for the estimator
psi_mom <- 154 / 162
lower_mom <- psi_mom - z * se
upper_mom <- psi_mom + z * se
```

95% Confidence Interval of Bootstrap Samples for MOM Method

We plot the 95% confidence intervals for the vaccine efficacy estimate over 100 bootstrap samples.Below is the R code to visualize the confidence intervals of 100 samples:

```r
# Create a dataframe for plotting
sample_summary <- data.frame(
  sample = 1:B,
  lower = lower_mom_boot,
  upper = upper_mom_boot,
  estimate = boot_values
)

# Plot the confidence intervals for the bootstrap samples
ggplot(data = sample_summary) +
  geom_segment(mapping = aes(x = lower, xend = upper, y = sample, yend = sample),
               color = "blue") +
  labs(x = "Sample Confidence Interval", y = "Bootstrap Samples",
       title = "90% Confidence Interval for 100 Bootstrap Samples") +
  theme_minimal() +
  theme(axis.text.y = element_blank(), axis.ticks.y = element_blank()) +
  geom_vline(xintercept = psi_mom, color = "red", linetype = "dashed") +
  annotate("text", x = psi_mom, y = B * 0.95, label = "Observed Efficacy",
           color = "red", angle = 90, vjust = -0.5)
```

Empirical P-value Calculation for MOM Method

We calculated the empirical P-value by simulation.

```r
# Calculate empirical p-value
set.seed(123)
B <- 100
psi_null = 0.3
n = 170
x = 8
### probability of COVID with vaccine
```

```r
cases_vaccine = (1-psi_null)/(2-psi_null)

### probability of COVID with no vaccine

cases_placebo = 1-cases_vaccine


for(i in 1:B) {

  sample_boot = sample(c(1, 0), size = n, replace = TRUE,

                       prob = c(cases_vaccine, cases_placebo))

  x = length(which(sample_boot == 1))

  boot_values[i] = (n - 2 * x) / (n - x)

}


empirical_p_value <- mean(boot_values >= psi_mom)
```