

Unmasking the Numbers

An Analysis of the BNT162b2 mRNA Covid-19 Vaccine Efficacy

Nathan Dennis, Aarav Dewangan, Maxwell Wang

1 Abstract

The COVID-19 (SARS-CoV-2) global pandemic's widespread health, economic, and social impact has necessitated the creation of effective vaccines, such as Pfizer's BNT162b2 vaccine. A placebo-controlled experiment with nearly 35,000 participants was conducted to determine the efficacy of the BNT162b2 vaccine candidate. The results from this study are analyzed using three approaches: Bayesian, Frequentist (likelihood), and bootstrapping. We analyze whether or not the vaccines efficacy is greater than 30%, required for FDA approval using statistics such as confidence intervals, point estimates, and hypothesis testing at a 5% level. It was discovered the BNT162b2 vaccine successfully had greater than a 30% vaccine efficacy using all three approaches, usually around 95% efficacy. We also concluded Pfizer used a rather uninformative prior in their own analysis.

Keywords

COVID-19, Vaccine Efficacy, Posterior, Likelihood, Simulation

2 Introduction

This paper focuses on the COVID-19 pandemic. There have been over 700 million confirmed cases of COVID-19 since the pandemic was identified by the World Health Organization on March 11, 2020. The symptoms range from mild flu-like symptoms to more severe respiratory issues. This pandemic has led to the loss of 7 million human lives worldwide and has caused many challenges to public health, food systems, and the work force, to name a few (World Health Organization). The

effect on the economy has been equally as devastating, with nearly half of the worlds' 3.3 billion global workforce at risk of losing their livelihoods (World Health Organization).

Scientists and medical professionals have been working diligently to develop and test vaccines in hopes to counteract this virus. One of these vaccines was developed by Pfizer and BioNTech in December 2020 when they successfully obtained a US FDA Emergency Use Authorization, where they began to distribution their two-dose vaccine, BNT162b2. These vaccines must go through a testing trial, where randomly assigned participants are assigned to a treatment/vaccine group and a control/placebo group. Such analysis may be compared to Sinovac Research & Development Co., Ltd's own COVID-19 vaccination efficacy trial, which used a similar approach through randomization and outcome (Sinovac). The point of these experiments are to test the vaccine efficacy, which is based on how many people who got vaccinated developed COVID-19 compared with how many people who got the placebo developed COVID-19 to determine whether the drug's intervention created an impact on the patients outcome (Centers for Disease Control). These trials are important to make decisions about the effectiveness of medicine.

Turning the focus to the BNT162b2 vaccine, the Clinical Trial Group ran a placebo-controlled, observer blinded efficacy study where randomly assigned persons 16 years or older in a 1:1 ratio received two doses, 21 days apart, of either the placebo or BNT162b2 vaccine. The key outcome of interest is the efficacy of the BNT162b2 vaccine, which we will test using a Bayesian, Frequentist, and Simulation approach. We are interested if the efficacy of the vaccine is greater than 30%, required for FDA approval.

3 Statistical Methods

3.1 Definitions & Initial Calculations

In this section, we need to make some calculations to create our statistical model. Let X denote the total number of laboratory confirmed Covid-19 cases among $n_1 = 17411$ subjects randomly assigned to the BNT162b2 vaccine group and let Y denote the total number of laboratory confirmed Covid-19 cases among $n_2 = 17511$ subjects randomly assigned to the placebo group. Also, let parameters π_p represent the probability of being infected with Covid-19 for an individual in the placebo group,

and π_v to represent the probability of being infected with Covid-19 for an individual in the vaccine group. It is noteworthy that X is independent of Y , and there are $n_1 + n_2$ total subjects

$$X \sim \text{Binom}(n_1, \pi_v) \qquad Y \sim \text{Binom}(n_2, \pi_p)$$

We have a total of $X + Y = s$ COVID-19 cases observed where now let T denote the number of individuals who were in the BNT162b2 vaccine group from the $s = 170$ patients who got COVID-19. T then follows the distribution: $T \sim \text{Binom}(s, \pi)$ (Lemma 1, Appendix) with success probability π , representing the probability of being in the vaccine group given an individual has COVID-19.

First, define the following variables for simplicity: Let V : A randomly assigned persons is in the vaccine group. Let C : A randomly assigned persons is in the placebo group. Let I : A laboratory confirmed COVID-19 infection. Also, recall the distribution of T well approximated by binomial for the BNT162b2 group study. Therefore: $T \sim \text{Binom}(s, \pi)$, where $\pi = P(V|I)$

Using Bayes Theorem, we can define π as $\pi = \frac{P(I|V)P(V)}{P(I)} = \frac{P(I|V)P(V)}{P(I|V)P(V) + P(I|C)P(C)}$. Each of these probabilities can be solved for using results defined previously, proved in Lemma 2 in the Appendix.

Calculation of π : Given these probabilities, we can now calculate $\pi = \frac{\pi_v \frac{n_1}{n_1+n_2}}{\frac{\pi_v n_1}{n_1+n_2} + \frac{\pi_p n_2}{n_1+n_2}} = \frac{\pi_v n_1}{\pi_v n_1 + \pi_p n_2}$.

When $n_1 \approx n_2$, we say randomization is 1:1 and are able to simplify π further: $\pi = \frac{\pi_v}{\pi_v + \pi_p}$

From here, the parameter of interest is the vaccine efficacy ψ , which is written as $\psi = \frac{\pi_p - \pi_v}{\pi_p}$, estimated by $\hat{\psi} = \frac{\bar{Y} - \bar{X}}{\bar{Y}}$. Given $\pi = \frac{\pi_v}{\pi_v + \pi_p}$, rearranging yields $\pi\pi_v + \pi\pi_p = \pi_v$. Furthermore, by grouping π_v the result is as follows: $\pi_v(1 - \pi) = \pi\pi_p$, then $\pi_v = \frac{\pi\pi_p}{1-\pi}$.

Calculation of ψ : Plug this into $\psi = \frac{\pi_p - \pi_v}{\pi_p} = 1 - \frac{\pi_v}{\pi_p}$, we get that $\psi = 1 - \frac{\pi_v}{\pi_p} = 1 - \frac{\pi\pi_p}{(1-\pi)\pi_p}$. Dividing by π_p on the right, $\psi = 1 - \frac{\pi}{1-\pi} = \frac{1-2\pi}{1-\pi}$. So, $\psi = \frac{1-2\pi}{1-\pi}$.

We now can construct the hypothesis of interest for ψ . The FDA requires at least 30% efficacy for a new therapy to be approved. The null hypothesis is $H_0 : \psi \leq 0.3$, and the alternative is $H_1 : \psi > 0.3$. Under these conditions $H_0 : \psi = 0.3$ implies $H_0 : \psi \leq 0.3$. We use a significance level of 0.05, 5%.

Within this section the following rearrangement of terms is used, given $\psi = \frac{1-2\pi}{1-\pi}$, we rearranged to get $\pi = \frac{1-\psi}{2-\psi}$. We can proceed with the three approaches to analyze the data.

3.2 Bayesian Approach

(1) Prior from Polack et. al.

Using results we previously defined, we have $T \sim Binom(170, \pi_0)$ where $\pi = \frac{\pi_v}{\pi_v + \pi_p}$. We will also use the prior distribution for π , $g(\pi_0) = Beta(0.700102, 1)$ as reported by Polack et. al. In the article, it was also mentioned that the 95% confidence interval for π is (0.005148, 0.964483) and the corresponding 95% interval for ψ is (-26.2, 0.995). This implies we have enough room for uncertainty.

We know beta is a conjugate prior in the binomial model. In the other words, posterior distribution of π is also a beta distribution, namely $h(\pi_0|x) = Beta(t + \alpha_0, 170 - t + \beta_0)$ given $g(\pi_0) = Beta(\alpha_0, \beta_0)$ in the binomial model, where $t=8$ in the dataset where there were 8 individuals in the vaccine group over all 170 cases. (Theorem 1) Using this fact to calculate the posterior:

$$h(\pi_0|t) = Beta(0.700102 + 8, 170 - 8 + 1) = h(\pi_0|t) = Beta(8.700102, 163)$$

Using this posterior distribution we can find the 95% Bayesian credible interval for π_0 , where we transform this for the vaccine efficacy, $\psi_0 = \frac{1-2\pi_0}{1-\pi_0}$, and calculate a p-value to test the null hypothesis, $P(\psi_0 < 0.3) = P(\pi > \frac{0.7}{1.7})$. We use a 95% interval similar to that used in the previous study. Building off the credible interval we can also calculate the posterior median for π in order to obtain the posterior median for ψ , finding the 0.5 percentile for the posterior distribution for π to do so.

(2) Alternative Beta Prior, Further Pessimistic Approach

Polack et. al. cites a prior distribution for π of $Beta(0.700102, 1)$ which is intended to be centered at $\pi = 0.4118$ corresponding to a ψ of 30%. The authors cite this as a pessimistic estimate. However, we will try a more pessimistic prior, where the median percentile of vaccine efficacy rate is 0% and 95th percentile is 0.2. A “pessimistic” prior would represent a more skeptical approach in regards to vaccine efficacy. It is argued that pessimistic priors are rarely necessary since researchers are motivated to conduct a trial with belief that the vaccine will be beneficial, however this is a flawed argument since in many cases prior beliefs about some intervention could prove to be inaccurate, causing more harm. (Zampieri, Fernando G, et al.) Due to this, using this prior is a safe option to eliminate the possibility of substantial harm and would be useful for comparison

We calculate $P(\psi < 0) = 0.5 \implies P(\frac{1-2\pi}{1-\pi} < 0) = 0.5 \implies P(\pi > \frac{1}{2}) = 0.5$ and $P(\psi < 0.3) =$

$0.95 \implies P(\frac{1-2\pi}{1-\pi} < 0.2) = 0.95 \implies P(\pi > \frac{0.8}{1.8}) = 0.95$. Using software one can observe the quantiles which correspond to this value within a beta distributon are $\alpha = \beta = 109.29$. This corresponds to a prior of $g(\pi_0) = \text{Beta}(109.29, 109.29)$. Using Theorem 1, we see that the posterior takes the form: $h(\pi_0|t) = \text{Beta}(109.29 + 8, 170 - 8 + 109.29) = \text{Beta}(117.29, 271.29)$.

Using this posterior distribution we can find the 95% Bayesian credible interval for π_0 , where we transform this for the vaccine efficacy, $\psi_0 = \frac{1-2\pi_0}{1-\pi_0}$ and calculate a p-value to test the null hypothesis. We use a 95% interval similar to that used in the previous study. Building off the credible interval we can also calculate the posterior median for π in order to obtain the posterior median for ψ .

(3) Poisson Distribution - Gamma Prior (Jeffery's Prior - Non Informative Prior)

So far we have constructed beta priors and posteriors from the Binomial distribution. Suppose now we can construct a Poisson distribution using the previous binomial distribution defined as $T \sim \text{Binom}(170, \pi_0)$. A good approximation for the binomial when the number of samples is large and π is small would be a poisson distribution with $\lambda = 170\pi_0$. We can denote $U \sim \text{Poisson}(\lambda = 170\pi_0)$, where λ represents the rate of covid cases in the vaccine group over 170 positive cases and U represents the number of covid cases in the vaccine group over the 170 positive cases.

For a Poisson distribution, we use Jeffreys prior for the rate parameter $\lambda \geq 0$, which states that $p(\lambda) = \sqrt{\frac{1}{\lambda}}$. (Fan, Theorem 2). This can also be thought of as a $\text{Gamma}(\frac{1}{2}, 0)$ distribution, which implies the posterior takes the form $h(\lambda_0|u) \sim \text{Gamma}(\frac{1}{2} + u, 170)$ for the 170 individual cases. (Fan, Theorem 2) Using this posterior we create a 95% Bayesian credible interval for π_0 , then transform this for the vaccine efficacy, $\psi_0 = \frac{1-2\pi_0}{1-\pi_0}$, calculate the median and p-value to test the hypothesis.

3.3 Frequentist: Maximum-Likelihood Estimation

We want to make a maximum-likelihood estimator (MLE) for $\psi, \widehat{\psi_0^{mle}}$, given the observed data. Let us start by writing the likelihood function. We have shown previously that $T \sim \text{Binom}(170, \pi_0)$, where the distribution would be defined as: $f(t|\pi_0) = \binom{170}{t} \pi_0^t \cdot (1 - \pi_0)^{170-t}$, where there are 170 total positive COVID-19 cases. We need to reparamaterize this with ψ , which represents the number of patients who contracted COVID who had initially been in the vaccinated group rather than the placebo. We use the fact that $\pi = \frac{1-\psi}{2-\psi}$, so $f(t) = \binom{170}{t} (\frac{1-\psi_0}{2-\psi_0})^t \cdot (1 - \frac{1-\psi_0}{2-\psi_0})^{170-t}$, $t = 0, 1, \dots, s$

Plugging in this value for π , we now have that $L(\psi) = f(t|\psi) = \binom{170}{t} \left(\frac{1-\psi}{2-\psi}\right)^t \cdot \left(1 - \frac{1-\psi}{2-\psi}\right)^{170-t}$. So the likelihood function is as follows:

$$L(\psi) = \binom{170}{t} \left(\frac{1-\psi}{2-\psi}\right)^t \cdot \left(1 - \frac{1-\psi}{2-\psi}\right)^{170-t} = \binom{170}{t} \left(\frac{1-\psi}{2-\psi}\right)^t \cdot \left(\frac{1}{2-\psi}\right)^{170-t} \quad -\infty < \psi \leq 1$$

Next, we compute the log-likelihood function by taking the natural log of the likelihood function.

We denote this log-likelihood as $\ell(\psi) = \ln(L(\psi)) = \ln(f(t|\psi)) = \ln\left(\binom{170}{t} \left(\frac{1-\psi}{2-\psi}\right)^t \cdot \left(\frac{1}{2-\psi}\right)^{170-t}\right)$.

$$\begin{aligned} \ell(\psi) &= \ln\left(\binom{170}{t} \left(\frac{1-\psi}{2-\psi}\right)^t \cdot \left(\frac{1}{2-\psi}\right)^{170-t}\right) \\ &= \ln\left(\binom{170}{t}\right) + \ln\left(\left(\frac{1-\psi}{2-\psi}\right)^t\right) + \ln\left(\left(\frac{1}{2-\psi}\right)^{170-t}\right) \\ &= \ln\left(\binom{170}{t}\right) + t \ln\left(\frac{1-\psi}{2-\psi}\right) + (170-t) \ln\left(\frac{1}{2-\psi}\right) \\ &= \ln\left(\binom{170}{t}\right) + t \ln(1-\psi) - t \ln(2-\psi) + (170-t) \ln(1) - (170-t) \ln(2-\psi) \\ &= \ln\left(\binom{170}{t}\right) + t \ln(1-\psi) - 170 \ln(2-\psi) \end{aligned}$$

We have calculated the log-likelihood $\ell(\psi) = \ln\left(\binom{170}{t}\right) + t \ln(1-\psi) - 170 \ln(2-\psi)$. To find the maximum of the likelihood function, we now take derivative of the log-likelihood and set it to zero. First, take the derivative with respect to ψ .

$$\frac{d}{d\psi} \ell(\psi) = \frac{d}{d\psi} \left(\ln\left(\binom{170}{t}\right) + t \ln(1-\psi) - 170 \ln(2-\psi) \right) = 0 - \frac{t}{1-\psi} + \frac{170}{2-\psi}$$

Using this derivative, we can set it equal to 0 to solve for the critical point ψ , MLE candidate:

$$\begin{aligned} 0 &= -\frac{t}{1-\psi} + \frac{170}{2-\psi} \\ \frac{t}{1-\psi} &= \frac{170}{2-\psi} \\ 2t - t\psi &= 170 - 170\psi \implies 2t - 170 = t\psi - 170\psi \\ \psi &= \frac{2t - 170}{t - 170} \end{aligned}$$

So, $\hat{\psi}_0^{mle} = \frac{2t-170}{t-170}$. In our case, we have that $t=8$, since T represents the number of covid cases in

the vaccine group within the 170 cases, we have given in the dataset (Table 1) $t=8$ and can plug this value in to solve for $\hat{\psi}_0^{mle}$ in the results. However, to prove that this critical point is a maximum we utilize the second derivative test.

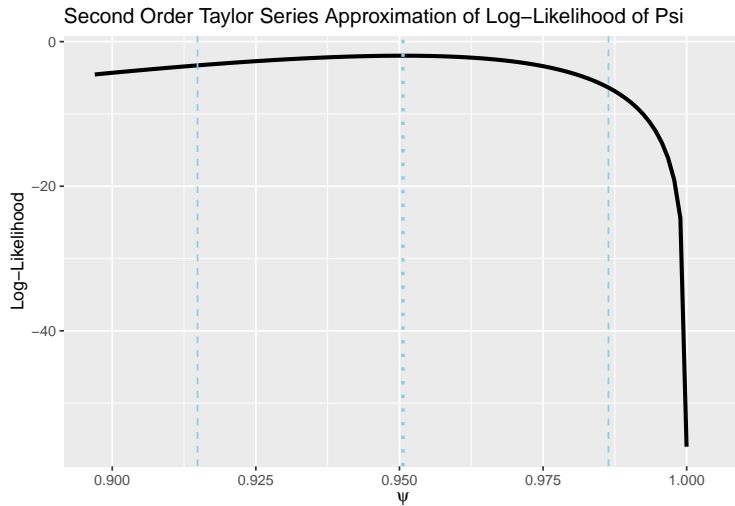
$$\frac{d^2}{d\psi^2}\ell(\psi) = \frac{d}{d\psi} \left(-\frac{t}{1-\psi} + \frac{170}{2-\psi} \right) = -\frac{t}{(1-\psi)^2} + \frac{170}{(2-\psi)^2}$$

From here we plug in the critical point, $\hat{\psi}_0^{mle}$, we found:

$$\begin{aligned} \frac{d^2}{d\psi^2}\ell(\psi) &= -\frac{t}{(1 - \frac{2t-170}{t-170})^2} + \frac{170}{(2 - \frac{2t-170}{t-170})^2} \implies -\frac{t}{(\frac{-t}{t-170})^2} + \frac{170}{(\frac{170}{t-170})^2} \\ &= -\frac{1}{t}(t-170)^2 + \frac{1}{170}(t-170)^2 \\ &= \left(\frac{1}{170} - \frac{1}{t} \right) (t-170)^2 \end{aligned}$$

We know that t cannot be a negative number and that $t \leq 170$ since there are 170 positive cases. So, the corresponding fraction $\frac{1}{170} - \frac{1}{t} \leq 0$ and the square term must be positive, so $\frac{d^2}{d\psi^2}\ell(\hat{\psi}_0^{mle}) < 0$, implying the critical point and hence MLE is a maximum.

We now have our estimate for ψ , where we can assess regulatory conditions. Suppose we now have that $T_1, T_2, \dots, T_{170} \sim \text{Bernoulli}(\pi)$, representing the 170 total infections within the study. Since the number of cases is large, under certain regulatory conditions $\hat{\psi}_0^{mle} \sim \text{Norm} \left(\psi, \frac{1}{\sqrt{-\ell''(\hat{\psi}_0^{mle})}} \right)$. We discuss the conditions as we observe a second order Taylor approximation of the log-likelihood function of ψ , where we know the first of which is passed.



The conditions we assess are that the possible values of our parameter and bounds don't depend on ψ , and that the true value of ψ does not fall on a boundary. We see in the plot that the second order Taylor series approximation for ψ is a very good fit and the estimated value for ψ isn't too close to a boundary. Thus, we have that the normality of the MLE for ψ holds. Next we want to construct a 95% Wald confidence interval for ψ , so we first compute the standard error which is defined to be $\hat{SE}(\hat{\psi}_0^{mle}) = \frac{1}{\sqrt{-\ell''(\hat{\psi}_0^{mle})}} = \frac{1}{\sqrt{(\frac{1}{170} - \frac{1}{t})(t-170)^2}}$, where again we know $t=8$ in the dataset from Table 1 and we can find this in the results. The Wald confidence interval is defined to take the form: ($z_{\alpha/2}$ is the $1 - \frac{\alpha}{2}$ th quantile of a standard normal)

$$\hat{\psi}_0^{mle} \pm z_{\alpha/2} \sqrt{\frac{1}{-\ell''(\hat{\psi}_0^{mle})}}$$

Finally, we utilize the likelihood ratio test for the hypothesis: $H_0 : \psi_0 = 0.3$ and $H_1 : \psi_0 \neq 0.3$. This is defined to be $W = 2 \ln \left[\frac{\hat{\psi}_0^{mle}}{\hat{\psi}_0^{null}} \right]$. Under the null, $W \approx \chi_1^2$. Given the equations provided previously, the statistic takes the form:

$$W = 2 \ln \left[\frac{\binom{170}{t} \left(\frac{1-\hat{\psi}_0^{mle}}{2-\hat{\psi}_0^{mle}} \right)^t \left(\frac{1}{2-\hat{\psi}_0^{mle}} \right)^{170-t}}{\binom{170}{t} \left(\frac{1-\hat{\psi}_0^{null}}{2-\hat{\psi}_0^{null}} \right)^t \left(\frac{1}{2-\hat{\psi}_0^{null}} \right)^{170-t}} \right]$$

Large values of W provide evidence in the direction of the null, so we observe the right-tail probability.

3.4 Bootstrap Confidence Interval

Using $\hat{\pi}_p$ and $\hat{\pi}_v$ from the data as the parameters for our binomial distributions X and Y (see section 3.1), we will run a parametric bootstrap simulation with $B = 10,000$ trials and construct a corresponding bootstrap confidence interval. Recall vaccine efficacy ψ is defined as $\frac{\pi_p - \pi_v}{\pi_p} = 1 - \frac{\pi_v}{\pi_p}$. So in terms of our bootstrap, $\psi \approx 1 - \frac{X/n}{Y/m}$ to estimate the efficacy of the vaccine.

The bootstrap simulation will take 10,000 samples of X and Y from $X \sim \text{Binom}(17411, \frac{8}{17411})$, $Y \sim \text{Binom}(17511, \frac{162}{17511})$, from the data in Table 1. This uses parametric bootstrapping from the assumed binomial distribution, with two defined outcomes, being infected with COVID-19 or not. We create re-samples to make inference for the parameter, ψ , using a confidence interval.

4 Results

First, we display the table, denoted Table 1, which corresponds to the data obtained in the BNT162b2 vaccines trials (Polack et al.).

Table 1: Vaccine Efficacy against Covid-19 at least 7 days after second dose in patients without evidence of infection

Groups	Cases	No. of Subjects
BNT162b2	8	17,411
Placebo	162	17,511
Total	170	34,922

Using Table 1, we observe that there is nearly a 1:1 ratio for number of subjects in the vaccine (17,411) and placebo group (17,511), which is expected. We can calculate the vaccine efficacy strictly based on the given results in the table. First, we have that $\hat{\pi}_v = \frac{8}{17411}$ which represents the probability of infection for someone in the vaccine group and $\hat{\pi}_p = \frac{162}{17511}$ which represents the probability of infection for someone in the placebo group. So, we calculate $\hat{\psi} = \frac{\hat{\pi}_p - \hat{\pi}_v}{\hat{\pi}_p} = \frac{\frac{162}{17511} - \frac{8}{17411}}{\frac{162}{17511}}$ this simplifies too: $\hat{\psi} = \frac{446749}{470097} = 0.950334$ as the observed vaccine efficacy, much larger than 0.3.

4.1 Prior 1: Beta(0.700102, 1)

We calculate the 95% credible interval for π_0 , [0.023194, 0.08799], and the posterior median as $\pi_0 = 0.048932$. We can convert this to create an interval for ψ using the fact that $\psi = \frac{1-2\pi}{1-\pi}$, plugging in the credible interval values. We obtained a 95% Bayesian credible interval reported as [0.903512, 0.976255].

We can also calculate a p-value for this to confirm our beliefs. As a reminder, $H_0 : \psi_0 \leq 0.3$ and $H_1 : \psi_0 > 0.3$, where the p-value was calculated to be 1.960014×10^{-28} , which is nearly 0.

Furthermore, we calculate the posterior median as $\pi_0 = 0.048932$. We then see that the posterior median for ψ comes out to be 0.948551, which is extremely large.

4.2 Prior 2: Beta(109.29, 109.29)

Using R software we calculate the 95% credible interval for π_0 to be [0.257257, 0.348361]. We again convert this to create an interval for ψ using the equality $\psi = \frac{1-2\pi}{1-\pi}$. So our 95% interval for ψ_0 is [0.465409, 0.65364].

Again we can also calculate a p-value under the hypothesis: $H_0 : \psi_0 \leq 0.3$ and $H_1 : \psi_0 > 0.3$, where the p-value was calculated to be 6.306123×10^{-8} , which is nearly 0.

We calculate the posterior median as $\pi_0 = 0.274467$. We the see that the posterior median for ψ_0 equals 0.621702, which is relatively large.

4.3 Prior 3 (Jeffery's Prior): Gamma(1/2, 0)

Using R software we calculate the 95% credible interval for π_0 to be [0.022248, 0.088797]. We can convert this to create an interval for ψ using the fact that $\psi = \frac{1-2\pi}{1-\pi}$, plugging in the credible interval values. We obtained a 95% Bayesian credible interval reported as [0.902549, 0.977246].

We can again calculate a p-value for this under $H_0 : \psi_0 \leq 0.3$ and $H_1 : \psi_0 > 0.3$, where the p-value was calculated to be 2.181923×10^{-21} , which is nearly 0.

We calculate the posterior median as $\pi_0 = 0.048053$. We the see that the posterior median for ψ_0 is then 0.949521, which is extremely large.

4.4 Table for Overall Bayesian Inference

Table 2: Bayesian Inference

Prior	Median for ψ_0	95% Credible Interval for ψ_0	p-value
Beta(0.700102, 1)	0.948551	[0.903512, 0.976255]	1.960014×10^{-28}
Beta(109.29, 109.29)	0.621702	[0.465409, 0.65364]	6.306123×10^{-8}
Gamma($\frac{1}{2}$, 0)	0.949521	[0.902549, 0.977246]	2.181923×10^{-21}

4.5 Maximum-Likelihood

Likelihood Calculations

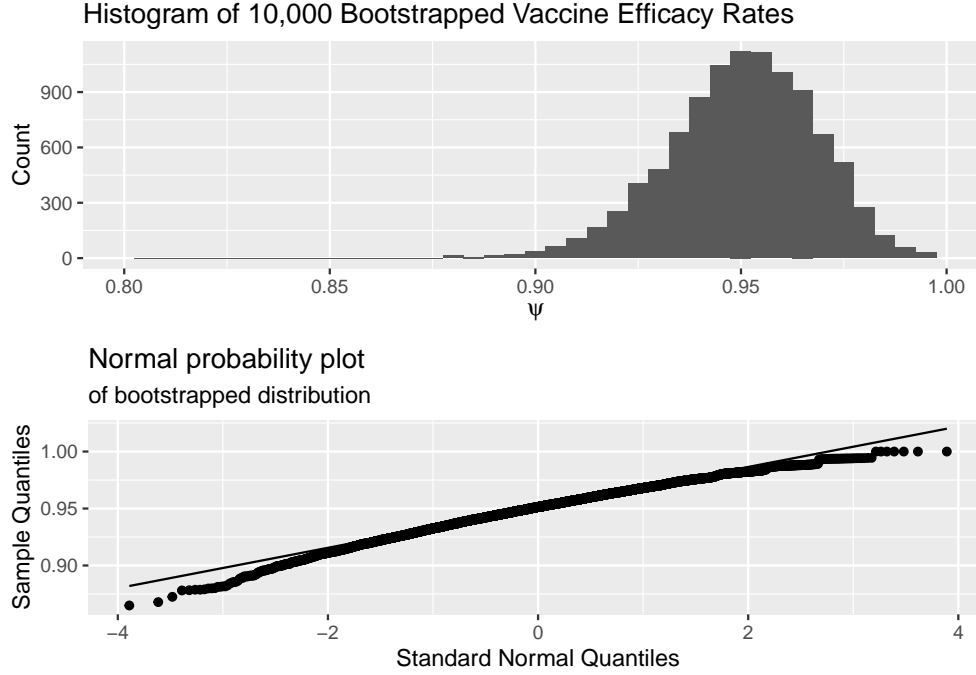
We use the derivation for the maximum likelihood for ψ , with t=8 from Table 1, to get: $\hat{\psi}_0^{mle} = \frac{170-2t}{170-t} = \frac{170-2 \cdot 8}{170-8} = \frac{77}{81}$. We showed that the second derivative yield a negative value, so this is a maximum.

We calculated the standard error for the MLE as $\frac{1}{\sqrt{(\frac{1}{170}-\frac{1}{8})(8-170)^2}} \approx 0.017885$. We then construct a 95% Wald Confidence interval, which was [0.915563, 0.985672] using $\frac{77}{81} \pm 1.96 \frac{1}{\sqrt{(\frac{1}{170}-\frac{1}{8})(8-170)^2}}$.

Furthermore, we can calculate the w-statistic from the likelihood ratio test where the equation was defined previously, which came out to be 121.6012 with a corresponding p-value of 2.82294×10^{-28}

4.6 Bootstrap Confidence Interval

The bootstrapped data was as follows. We check for normality:



From the code in our appendix, we calculated the following confidence intervals for vaccine efficacy ψ .

Table 3: Bootstrapped Confidence Intervals for ψ_0

	Lower	Upper
Percentile Method	0.912194	0.982040
Standard Bootstrap	0.915071	0.985596

Based on the Normal Q-Q plot, we conclude that the bootstrapped distribution is not nearly normal as the tail ends of the Q-Q plot differ from the expected values should our bootstrapped values of ψ_0 come from a normal distribution. Therefore, we should prefer the percentile method 95% confidence interval for ψ_0 instead.

5 Discussion/Conclusion

Using the Bayesian methods and 3 priors examined in Table 2, all 3 priors had 95% credible intervals for ψ which do not contain the value 0.3 and are completely above this value. The intervals for the first prior, used from the Pfizer study, and third prior, using the Gamma prior, were around the same, both above 90%. The second prior, using a very pessimistic approach, was smaller than these 2 credible intervals, but still above 0.3. We can conclude using these credible intervals that the vaccine efficacy is better than 30%. We also calculated p-values for all 3 prior approaches, but all p-values were nearly 0 with the hypothesis $H_0 : \psi_0 \leq 0.3$ and $H_1 : \psi_0 > 0.3$. So, since these p-value are below the significance level of 0.05 we have enough evidence to reject the null in favor of the alternative, concluding vaccine efficacy for the BNT162b2 vaccine is greater than 30%. Also observe that for each prior the median for ψ is significantly higher than 0.3, further evidence that $\psi > 0.3$. As stated, all 3 different methods produced the same conclusion. The extremely pessimistic prior produced results which were slightly different than the other two, but it could be stronger since the assumption about the prior is much more biased towards the vaccine efficacy rate smaller than 30%. It is noteworthy that Pfizers prior and Jeffery's prior, known to be non-informative, were very similar. If the Pfizer prior was overly optimistic, it would skew results towards higher efficacies.

Moving on to the Frequentist approach, we calculated $\hat{\psi}_0^{mle} = \frac{77}{81}$. We see that this value is very large, as the estimated vaccine efficacy rate is around 0.950617. The MLE allows us to estimate the parameters of our model which maximizes the likelihood function, we have a very large estimate. Furthermore, we calculated the 95% wald confidence interval. This helps provide an estimate for the range of our parameter, a range of values for the true parameter, which we saw was [0.915563, 0.985672] which is well above 0.3, so there is strong evidence the vaccine efficacy is greater than 0.3. We are 95% confident that the true parameter, ψ falls in the range [0.915563, 0.985672].

Furthermore, the W statistic calculated was 121.6012, where large values of W are known to provide evidence in favor of the alternative, in this case $\psi_0 \neq 0.3$. The p-value also was extremely small and basically 0, where we can safely reject the null in favor of the alternative.

Finally, our bootstrapped confidence interval using the percentile method for ψ found that we are 95% confident that the true vaccine efficacy is between [0.9110567, 0.9817138] (see Table 3), with

our median bootstrapped $\psi = 0.95145$. Our entire confidence interval lies above 30% so there is sufficient statistical evidence to conclude that the vaccine is sufficiently effective and safely reject the null hypothesis. However, since we used a parametric bootstrap, it could be harmful if the distribution is not binomial prompting further analysis, possibly non-parametric bootstrapping.

All three methods created similar conclusions and allows us to interpret the results from multiple views, potentially limiting bias. We also observed that Pfizers prior performed very closely to a known non-informative prior, prompting the conclusion that likewise, it was non-informative.

6 References

Centers for Disease Control. “Vaccine Effectiveness Studies.” Centers for Disease Control and Prevention, <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/effectiveness/how-they-work.html>

Polack et al. Safety and Efficacy of the BNT162B2 Mrna Covid-19 Vaccine | Nejm, Dec. 2020, <https://www.nejm.org/doi/full/10.1056/nejmoa2034577>

Fan, Zhou. Lecture 20 | Bayesian Analysis - Stanford University, Nov. 2016, <https://web.stanford.edu/class/stats200/Lecture20.pdf>

Sinovac. “Clinical Trial of Efficacy and Safety of Sinovac’s Adsorbed COVID-19 (Inactivated) Vaccine in Healthcare Professionals - Full Text View.” Clinical Trial of Efficacy and Safety of Sinovac’s Adsorbed COVID-19 (Inactivated) Vaccine in Healthcare Professionals - ClinicalTrials.Gov, July 2020, <https://clinicaltrials.gov/ct2/show/NCT04456595>.

World Health Organization. “Impact of Covid-19 on People’s Livelihoods, Their Health and Our Food Systems.” World Health Organization, 13 Oct. 2020, <https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people%27s-livelihoods-their-health-and-our-food-systems>

Zampieri, Fernando G, et al. “Using Bayesian Methods to Augment the Interpretation of Critical Care Trials. an Overview of Theory and Example Reanalysis of the Alveolar Recruitment for Acute Respiratory Distress Syndrome Trial.” American Journal of Respiratory and Critical Care Medicine, 1 Mar. 2021, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7924582/>

7 Appendix

Lemma 1: Distribution of T

If we let T denote the number of infected individuals from the 170 patients who got infected with COVID-19, then we use the fact that the sample sizes for the vaccine group, $n_1 = 17,411$ and placebo group, $n_2 = 17,511$ is large and the rates π_v and π_p are small, we can say that $X \approx Pois(17,411)$ and $Y \approx Pois(17,511)$. It follows $P(T = t) = P(X = t | X + Y = s) = \binom{s}{t} \pi^t (1 - \pi)^{s-t}$

Lemma 2:

For $P(V)$, dividing the total number of individuals in the vaccine group by the total in the study yields this result, $P(V) = \frac{n_1}{n_1 + n_2}$

For $P(I|V)$, π_v was defined to be the probability of an infection in the vaccine group, so $P(I|V) = \pi_v$

For $P(I)$, the probability of a confirmed COVID-19 infection is dependent on whether or not an individual comes from the placebo or vaccine group, in which case $P(I) = P(I \cap (V \cup C)) = P((I \cap V) \cup (I \cap C))$. Due to these events being disjoint, we can add the probabilities as $P((I \cap V) + P(I \cap C))$.

Using the chain rule of probability, $P(I) = P(I|V)P(V) + P(I|C)P(C)$. π_p was defined as the probability of an infection in the placebo group, so $P(I|C) = \pi_p$. Furthermore, $P(C)$ can be defined by $P(C) = \frac{n_2}{n_1 + n_2}$ where the number of individuals in the placebo group is divided by those in total.

Plugging these values in: $P(I) = \pi_v \frac{n_1}{n_1 + n_2} + \pi_p \frac{n_2}{n_1 + n_2} = \frac{\pi_v n_1 + \pi_p n_2}{n_1 + n_2}$

Theorem 1: Posterior of a Beta Proof

Suppose $X \sim Binom(n, \pi_0)$ and we assume π follows a Beta distribution with shape parameters α_0 and β_0 , $g(\pi_0) = \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)} \pi_0^{\alpha_0 - 1} (1 - \pi_0)^{\beta_0 - 1}$. Since X is binomial, we have that $P(X = x | \pi = \pi_0) = \binom{n}{x} \pi_0^x (1 - \pi_0)^{n-x}$. We show that the posterior distribution, $h(\pi_0 | x) \sim Beta(x + \alpha_0, n - x + \beta_0)$.

$$\begin{aligned} h(\pi_0 | x) &= \frac{P(X = x | \pi = \pi_0) g(\pi_0)}{\int_0^1 P(X = x | \pi = \pi_0) g(\pi_0)} \\ &= \frac{\binom{n}{x} \pi_0^x (1 - \pi_0)^{n-x} \Gamma(\alpha_0 + \beta_0) \pi_0^{\alpha_0 - 1} (1 - \pi_0)^{\beta_0 - 1}}{\Gamma(\alpha_0) \Gamma(\beta_0) \int_0^1 \binom{n}{x} \pi_0^x (1 - \pi_0)^{n-x} \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0) \Gamma(\beta_0)} \pi_0^{\alpha_0 - 1} (1 - \pi_0)^{\beta_0 - 1} d\pi_0} = \frac{\binom{n}{x} (1 - \pi_0)^{n-x+\beta_0-1} \pi_0^{\alpha_0-1+x}}{\int_0^1 \binom{n}{x} (1 - \pi_0)^{n-x+\beta_0-1} \pi_0^{\alpha_0-1+x} d\pi_0} \end{aligned}$$

By observing the terms in the numerator, this is recognizable as the PDF of a Beta distribution with parameter $\alpha = x + 1$ and $\beta = n - x + 1$. So, $h(\pi_0|x) \sim Beta(x + \alpha_0, n - x + \beta_0)$. We know this since $\int_0^1 u^{r-1}(1-u)^{s-1}du = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$, implying $\frac{\Gamma(\alpha_0+\beta_0+n)}{\Gamma(\alpha_0+x)\Gamma(\beta_0+n-x)}(1-\pi_0)^{n-x+\beta_0-1}\pi_0^{\alpha_0-1+x}$. ■

Theorem 2: Jefferies Prior (Fan)

Given $T \sim Binom(170, \pi)$, we know that when s is large (170) and π is small the binomial distribution is well approximated by a poisson distribution, $U \sim Pois(170\pi)$.

Jefferies Prior is defined to be $g(\lambda) \propto \sqrt{I(\lambda)}$ with $I(\lambda)$ being the Fishers Information matrix, where this can be evaluated as: $g(\lambda) \propto \sqrt{-E \left[\frac{d^2}{d\lambda^2} \ln(f(u|\lambda)) \right]}$ (Fan)

For the poisson distribution we calculate the likelihood give we observe n infections, we can write:

$$f(u|\lambda) = \frac{\lambda^u e^{-\lambda}}{u!}. \text{ Taking the log-likelihood: } \ln(L(\lambda)) = \frac{\lambda^u e^{-\lambda}}{u!}$$

$$\begin{aligned} \ln(L(\lambda)) &= \ln\left(\frac{\lambda^u e^{-\lambda}}{u!}\right) \\ &= u \ln(\lambda) - u - \ln(u!) \end{aligned}$$

We can take the first and second derivative of this with respect to λ . First derivative: $\frac{d}{d\lambda}(u \ln(\lambda) - u - \ln(u!)) = \frac{u}{\lambda}$. Second derivative: $\frac{d}{d\lambda}\left(\frac{u}{\lambda}\right) = -\frac{u}{\lambda^2}$. If we take the expected value of this, it results in the final result:

$$\begin{aligned} E \left[\frac{d^2}{d\lambda^2} \ln(f(u|\lambda)) \right] &= E\left[-\frac{u}{\lambda^2}\right] \\ &= -\frac{1}{\lambda^2} E[U] \\ &= -\frac{\lambda}{\lambda^2} \quad \text{Expected value of a poisson is the rate parameter} \\ &= -\frac{1}{\lambda} \end{aligned}$$

Since we have that $g(\lambda) \propto \sqrt{-E \left[\frac{d^2}{d\lambda^2} \ln(f(u|\lambda)) \right]}$, we plug in the value calculated and see that $g(\lambda) \propto \sqrt{\frac{1}{\lambda}}$, which is often expressed as the improper $Gamma(\frac{1}{2}, 0)$ distribution. (Fan)

For find the general form of the posterior that is applicable in any context, we will define a poisson distribution with rate parameter λ , $X_i \sim Pois(\lambda)$ with n , iid, X_i 's. We also define

$$g(\lambda) = \text{Gamma}(\alpha, \beta)$$

$$h(\lambda|x) = \frac{\frac{e^{-n\lambda}\lambda^{n\bar{x}}}{\prod_{i=1}^n (x_i)!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}}{\int_0^\infty \frac{e^{-n\lambda}\lambda^{n\bar{x}}}{\prod_{i=1}^n (x_i)!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda}$$

After canceling constants we are left with $h(\lambda|x) \propto e^{-n\lambda}\lambda^{n\bar{x}}\lambda^{\alpha-1}e^{-\beta\lambda} = e^{-n\lambda+\beta\lambda}\lambda^{n\bar{x}+\alpha-1}$. Which follows $\text{Gamma}(n\bar{x} + \alpha, n + \beta)$. We can now use this posterior and plug in our own values. We know that the prior follows a $\text{Gamma}(\frac{1}{2}, 0)$ distribution, so $\text{Gamma}(n\bar{x} + \frac{1}{2}, n + 0)$. We have that $s = 170$, so there are 170 total positive cases for n . Then there are 8 individuals in the vaccine group within these 170 positive cases, so $n\bar{x} = 8$ and the posterior is defined to be $\text{Gamma}(8 + \frac{1}{2}, 170)$ ■

Prior & Posterior Medians

Let ϕ represent the posterior median for π , then all posterior medians for ψ_0 will be found by:

$$\begin{aligned} P(\pi < \phi) &= 0.5 \\ P(\frac{1-\psi}{2-\psi} < \phi) &= 0.5 \\ P(1-\psi < \phi(2-\psi)) &= 0.5 \\ P(1-2\phi < \psi - \psi\phi) &= 0.5 \\ P(\frac{1-2\phi}{1-\phi} < \psi) &= 0.5 \end{aligned}$$

To test the hypothesis we have that

$$\begin{aligned} P(\psi_0 < 0.3) &= P(\frac{1-2\pi_0}{1-\pi_0} < 0.3) \\ &= P(1-2\pi_0 < 0.3(1-\pi_0)) \\ &= P(1-0.3 < 2\pi_0 - 0.3\pi_0) \\ &= P(\frac{0.7}{1.7} < \pi_0) \end{aligned}$$

Prior 1 Calculations

```
# 95% Credible interval for pi
pi_ci <- qbeta(p=c(0.025,0.975), 8.700102, 163)

# 95% Credible interval for psi
upper_ci_prior1 <- (1-2*pi_ci[1])/(1-pi_ci[1])
lower_ci_prior1 <- (1-2*pi_ci[2])/(1-pi_ci[2])

# Posterior Median for pi & psi
postmedian1 <- qbeta(0.5, 8.700102, 163)
postmedian1_psi <- (1-2*postmedian1)/(1-postmedian1)

# Posterior p-value for psi
postpval1 <- pbeta(0.7/1.7, 8.700102, 163, lower.tail=F)
```

Prior 2 Calculations

```
quantile1 = list(p = 0.5, x = 0.50)
quantile2 = list(p = 0.05, x = 0.8/1.8)
params <- beta.select(quantile1,quantile2)

pi_ci2 <- qbeta(p=c(0.025,0.975), 109.29+8, 109.29+162)

# 95% Credible interval for psi
upper_ci_prior2 <- (1-2*pi_ci2[1])/(1-pi_ci2[1])
lower_ci_prior2 <- (1-2*pi_ci2[2])/(1-pi_ci2[2])

# Posterior Median for pi & psi
postmedian2 <- qbeta(0.5, 86.04+8, 86.04+162)
```

```

postmedian2_psi <- (1-2*postmedian2)/(1-postmedian2)

# Posterior p-value for psi
postpval2 <- pbeta(0.7/1.7, 86.04+8, 86.04+162, lower.tail=F)

```

Prior 3 Calculations

```

# Upper and lower bounds for pi credible interval
lower_gamma <- qgamma(0.025, 1/2+8, 170)
upper_gamma <- qgamma(0.975, 1/2+8, 170)

# 95% Credible Interval for psi
upper_ci_prior3 <- (1-2*lower_gamma)/(1-lower_gamma)
lower_ci_prior3 <- (1-2*upper_gamma)/(1-upper_gamma)

# Posterior Median for pi & psi
postmedian1 <- qgamma(0.5, 1/2+8, 170)
postmedian1_psi <- (1-2*postmedian1)/(1-postmedian1)

# Posterior p-value for psi
postpval1 <- pgamma(0.7/1.7, 1/2+8, 170, lower.tail=F)

```

Likelihood Calculations

```

t <- 8 #Observed 8 COVID-19 cases in vaccine group over total number of cases

# Calculate MLE Given equation
mle_psi <- (2*t-170)/(t-170)

# Calculate estimated standard error of Psi given equation
psi_se <- 1/sqrt((1/t-1/170)*(t-170)^2)

# 95% Wald Confidence Interval

```

```

lower_ci_wald <- mle_psi - qnorm(0.975) * psi_se
upper_ci_wald <- mle_psi + qnorm(0.95) * psi_se

# W statistic, likelihood ratio test
w_obs <- 2 * log((((1-mle_psi)/(2-mle_psi))^8 *
                (1/(2-mle_psi))^162)/
                (((1-0.3)/(2-0.3))^8 * (1/(2-0.3))^162))

#Likelihood Ratio test p-value
p_value <- pchisq(w_obs, df=1, lower.tail=F)

```

Regulatory Conditions for Likelihood

```

# Log-likelihood function
loglik.psi <- function(psi){
  ifelse(psi > 1, NA,
        log(choose(170,8)+8*log(1-psi)-170*log(2-psi)))
}

# Second order taylor approximation of the log-likelihood
second_order <- maxLik2(loglik=loglik.psi,
                       start=mle_psi,
                       method="NR")

# Plot this taylor approximation
ts_plot <- plot(second_order) %>%
  gf_labs(title="Second Order Taylor Series Approximation of Log-Likelihood of
           Psi", x=expression(psi), y="Log-Likelihood")

```

Bootstrap Confidence Interval

```
# Requirements
set.seed(342)

# Data from study
n <- 17411
m <- 17511
pi_v_hat <- 8/n
pi_p_hat <- 162/m

# Helper function to compute psi from bootstrapped data
get_efficacy <- function(x_bar, y_bar) {
  1 - (x_bar / y_bar)
}

# Number of trials, B
B <- 10000

# Vaccine efficacy psi for our dataset
psi <- get_efficacy(pi_v_hat, pi_p_hat)

# Run bootstrap
boot_sim <- lapply(1:B, FUN = function(i){
  # generate a re-sample from known distribution
  x_bar = rbinom(1, n, pi_v_hat) / n
  y_bar = rbinom(1, m, pi_p_hat) / m

  #calculate bootstrapped efficacy and return value as dataframe
  psi_bootstrapped = get_efficacy(x_bar, y_bar)
  data.frame(psi_bootstrapped)
```

```

})

boot_sim_psi <- do.call(rbind, boot_sim)

# Large histogram
p1 <- ggplot(data=boot_sim_psi,
             mapping = aes(psi_bootstrapped))+
  geom_histogram(binwidth =0.01)+
  labs(title="Histogram of 10,000 Bootstrapped Vaccine Efficacy Rates",x="VE",
       y="Count")+xlim(0, 1)

# Detailed histogram with normality
p2_1 <- ggplot(data=boot_sim_psi,
               mapping = aes(psi_bootstrapped))+
  geom_histogram(binwidth =0.005)+
  labs(title="Histogram of 10,000 Bootstrapped Vaccine Efficacy Rates",x=expression(psi),
       y="Count")+xlim(0.8, 1)

# Normal qq
p2_2 <- ggplot(data = boot_sim_psi,
               mapping = aes(sample = psi_bootstrapped)) +
  stat_qq(distribution = qnorm) +
  stat_qq_line(distribution = qnorm) +
  labs(title = "Normal probability plot",
       subtitle = "of bootstrapped distribution",
       x="Standard Normal Quantiles",
       y="Sample Quantiles")

# Center of bootstrapped confidence interval for psi

```

```

bootstrapped_center <- mean(boot_sim_psi$psi_bootstrapped)

# Compute Bootstrap confidence interval
# Percentile method
percentile_method<-boot_sim_psi %>%
  summarize(lower = quantile(psi_bootstrapped,.025),
    upper = quantile(psi_bootstrapped, .975))
percentile_method # See output

median <- quantile(boot_sim_psi$psi_bootstrapped, 0.5)

# Bias adjusted Standard bootstrap method
standard_bootstrap<-boot_sim_psi %>% summarise(boot_mean=mean(psi_bootstrapped),
  boot_se = sd(psi_bootstrapped),
  bias = boot_mean-psi,
  lower = psi + qnorm(p=0.025) * boot_se,
  upper = psi - qnorm(p=0.025) * boot_se)

standard_bootstrap # See output

summary_table <- matrix(c(percentile_method$lower,percentile_method$upper,
  standard_bootstrap$lower,standard_bootstrap$upper),
  ncol=2,byrow=TRUE)
colnames(summary_table) <- c("Lower","Upper")
rownames(summary_table) <- c("Percentile Method","Standard Bootstrap")
summary_table <- as.table(summary_table)

```