

Reanalyzing Pfizer’s BNT162b2 Vaccine Efficacy Using Bayesian and Frequentist Approaches

Isaac Muhlestein, Jian Kang, Troy Russo

3/16/2024

Abstract

The COVID-19 pandemic emphasized the need for vaccine development, with Pfizer’s BNT162b2 mRNA vaccine demonstrating 95% efficacy in clinical trials. This study reanalyzes the trial data using frequentist and Bayesian frameworks to estimate vaccine efficacy and evaluate hypotheses. Frequentist methods employed maximum likelihood estimation (MLE) and parametric bootstrap confidence intervals, while Bayesian analysis utilized Beta priors to model vaccine efficacy. Both approaches yielded vaccine efficacy estimates exceeding 94.8%, with Bayesian credible intervals (e.g., 90.3–97.6% under the original prior) slightly wider than frequentist confidence intervals (94.8–95.3%). Likelihood ratio and posterior probability tests overwhelmingly rejected the null hypothesis (vaccine efficacy $\leq 30\%$) with p -values < 0.0001 . Results confirm BNT162b2’s high efficacy and highlight methodological nuances: Bayesian intervals incorporated prior evidence, whereas frequentist methods provided precise, assumption-driven estimates. Despite differing uncertainty quantification, both frameworks strongly supported the vaccine’s effectiveness, reinforcing confidence in the vaccine.

Keywords

Maximum Likelihood Estimation, Bayesian Beta-Binomial Model, Parametric Bootstrap, Likelihood Ratio Test

Introduction / Background

The COVID-19 pandemic, caused by the coronavirus SARS-CoV-2, has had a massive global impact since it first emerged in late 2019, affecting health, economies, and daily life worldwide. Developing safe and effective vaccines quickly became crucial to help curb the spread and reduce severe illness. One of the earliest breakthroughs was the mRNA-based BNT162b2 vaccine from Pfizer and BioNTech, which showed impressive efficacy in clinical trials (Polack et al., 2021). By December 2020, the U.S. Food and Drug Administration (FDA) granted Emergency Use Authorization (EUA) for BNT162b2 (FDA, 2020)—a major turning point in fighting the pandemic. This authorization was supported by data from a phase 2/3 randomized, placebo-controlled trial involving over 34,000 participants, which reported a 95% vaccine efficacy (VE) against symptomatic COVID-19 (Polack et al., 2021).

Vaccine efficacy trials rely on rigorous statistical frameworks to evaluate the reduction in disease risk between vaccinated and unvaccinated groups. Traditional frequentist methods, such as confidence intervals and p-values, have long dominated clinical trial analyses. However, Bayesian approaches are increasingly employed for their ability to incorporate prior evidence and provide probabilistic interpretations of efficacy (Berry et al., 2006). The original BNT162b2 trial used a Bayesian beta-binomial model with a Beta(0.700102, 1) prior to estimate VE, defined as

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

where π represents the probability of infection in the vaccine group relative to placebo (Polack, et al., 2021). This analysis yielded a posterior median VE of 95% (95% credible interval: 90.3–97.6%) and a $>99.99\%$ posterior probability that VE exceeded 30%. These results aligned with findings from other mRNA vaccines like Moderna’s mRNA-1273 (Baden et al., 2021).

Despite the strong evidence presented by the original Bayesian analysis, comparing Bayesian and frequentist methodologies offers unique insights. Frequentist methods, which avoid prior distributions and focus on long-run frequency, remain the regulatory standard for many trials (FDA, 2020). A double analysis could strengthen our confidence in the results by examining similarities and differences between the two frameworks, particularly in estimating uncertainty intervals. For instance,

while Bayesian credible intervals reflect posterior probability, frequentist confidence intervals rely on repeated sampling interpretations—a distinction with practical implications for interpretation (Wasserstein et al., 2019).

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

Group	Cases	Sample Size
BNT162b2	8	17,411
Placebo	162	17,511
Total	170	34,922

The present study reanalyzes these data Table 1 using both Bayesian and frequentist approaches. The primary goals are to: (1) estimate VE and associated uncertainty under each framework, (2) compare the resulting efficacy point and interval estimates, and (3) evaluate the posterior probability (Bayesian) and p-values (frequentist) for the hypothesis that VE exceeds 30%. Our hypotheses are as follows:

Hypotheses:

- Both frameworks will confirm high VE ($>90\%$) for BNT162b2
- Uncertainty intervals will differ modestly between methods
- Both analyses will reject the null hypothesis ($VE \leq 30\%$) with overwhelming evidence

This comparative analysis not only reinforces confidence in BNT162b2’s efficacy but also highlights the complementary strengths of Bayesian and frequentist statistics in public health research.

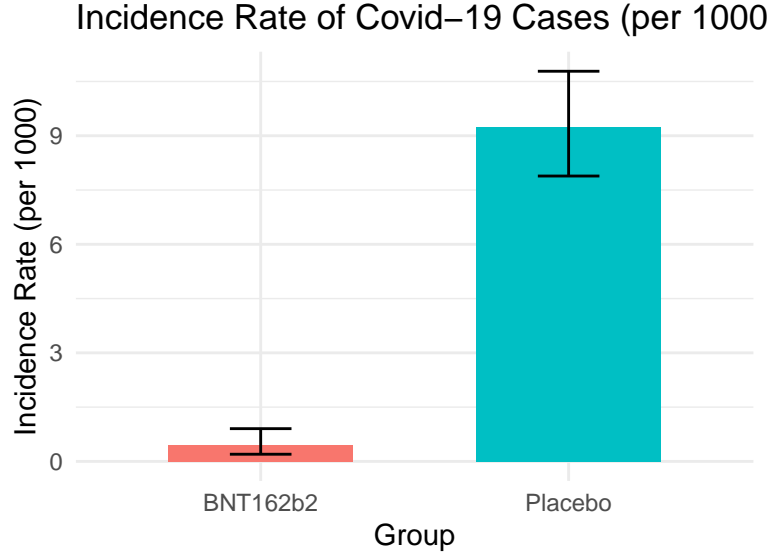


Table 2: Estimated Vaccine Efficacy (VE) based on observed attack rates.

Group	Attack Rate (%)
BNT162b2 (Vaccine)	0.046
Placebo	0.925
Vaccine Efficacy	95.03%

Statistical Methods

Model

The experiment uses a binomial model targeting 170 infections. π , the probability measure of this model, estimates the proportion of infected individuals that had received the vaccination. Because we are using a Binomial method, we are making assumptions that should be checked. We target 170 infections, so the constant trials assumption is met. Either the infected person received the vaccine or the placebo, so the assumption of only two outcomes is met. The parameter we are estimating, π , is assumed to have some unknown true value, an assumption we cannot strictly verify, but is reasonable to believe. Lastly, we assume that our observations are independent, which

we once again cannot strictly verify, but given our large sample size and the number of individuals involved in the survey, along with the fact that each participant's group was chosen randomly, this is a reasonable assumption. Thus, taking X to be the number of infections out of 170 where the patient is vaccinated, our experimental model is:

$$X \sim \text{Bin}(170, \pi)$$

However, the parameter that we are more interested in making inference on is ψ , which represents the efficacy of the vaccine. This parameter is calculated as

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

Notably, the FDA (Food and Drug Administration) requires that in order for a drug to be released for wholesale production and use, it must be shown to have an efficacy greater than 0.3. Thus, our hypothesis tests will test how statistically significant the evidence is that the vaccine has an efficacy greater than 0.3.

$$H_0 : \psi = 0.3; \quad H_1 : \psi \neq 0.3$$

We use two statistical methods to make this analysis: the frequentist maximum likelihood method, and the bayesian inference method.

Likelihood Inference

We can begin with the likelihood function. Recall that the design of our experiment is binomial; we observe 170 infections and try to make inference on π , the proportion of infections from those who have the vaccine. Assuming each observation is independent (which we have justified above), we can write our likelihood function in terms of π :

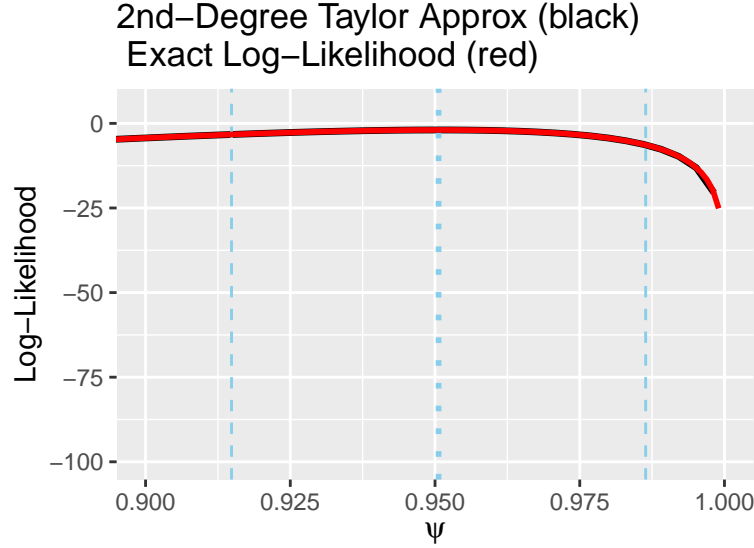
$$L(\pi) = \binom{170}{X} (\pi)^X (1 - \pi)^{170-X}$$

Where X is the observed number of infections where the patient had the vaccine.

We can rewrite this likelihood function in terms of ψ :

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

We then find the log-likelihood function. Pictured below is the second degree Taylor approximation of the log-likelihood:



To find the estimate of ψ that maximizes this likelihood function, we took the first derivative and maximized. The whole derivation, from likelihood function to final estimate formula, can be found in the Appendix. Our final formula is given by:

$$\hat{\psi}_0^{mle} = \frac{n - 2X}{n - X}$$

Given our estimate, we now perform a likelihood ratio test to test our hypotheses. We can initially find the likelihood ratio, Λ . It has formula given by:

$$\Lambda = \frac{L(\hat{\psi}_0^{mle})}{L(\psi_0^{null})}$$

Once we have calculated Λ , we can go forward with calculating W , which has formula given by:

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

W is especially helpful for making inference, because under the null hypothesis, $W \sim \chi_1^2$. Thus, we can use the value of W to find a p-value to evaluate our hypothesis, based on where it falls in the distribution of χ_1^2 (we will reject the null hypothesis for large values of W). For reference, a plot of the chi-squared distribution with 1 degree of freedom is given in the Appendix, marked with the boundary value beyond which we reject our null hypothesis.

The asymptotic variance of $\hat{\psi}^{\text{MLE}}$ is derived using the Fisher information of $\hat{\psi}$:

$$I(\psi) = \frac{n}{(1-\psi)(2-\psi)^2},$$

where n is the sample size. We construct the Wald confidence interval as follows:

$$\hat{\psi}^{\text{MLE}} \pm z_{\alpha/2} \cdot \frac{1}{\sqrt{nI(\hat{\psi}^{\text{MLE}})}}.$$

This interval assumes large-sample normality of the MLE. We will derive a 95% confidence interval using our maximum likelihood estimate (MLE).

Bootstrap Confidence Interval for $\hat{\psi}_{\text{MLE}}$

In addition to the large-sample interval, we use a computational bootstrap approach to estimate a 95% confidence interval for $\hat{\psi}_{\text{MLE}}$. Suppose we observe $\hat{\pi} = \frac{T}{n}$ from the sample, where T is the observed number of infections out of n participants. Then:

1. **Generate Resamples.** We repeatedly sample T^* from a Binomial($n, \hat{\pi}$) distribution. For each bootstrap draw:

$$\hat{\pi}^* = \frac{T^*}{n}, \quad \hat{\psi}^* = \frac{1 - 2\hat{\pi}^*}{1 - \hat{\pi}^*}.$$

2. **Form the Confidence Interval.** After repeating this process B times, let $\hat{\psi}_1^*, \hat{\psi}_2^*, \dots, \hat{\psi}_B^*$ be the resulting bootstrap estimates of $\hat{\psi}$. We then take the 2.5th and 97.5th percentiles of the $\hat{\psi}^*$ values to form the approximate 95% confidence interval:

$$[\hat{\psi}_{(0.025)}^*, \hat{\psi}_{(0.975)}^*].$$

This bootstrap procedure avoids strong distributional assumptions and can provide more reliable interval estimates for moderate sample sizes.

Bayesian Inference

We can use a Bayesian approach to reanalyze the efficacy of the BNT162b2 vaccine. Let T be the number of infections observed among n vaccine recipients in the trial. We assume

$$T \mid \pi \sim \text{Binomial}(n, \pi)$$

where π is the probability of infection in the vaccine group. As in the original trial design, vaccine efficacy (VE) is defined by Equation (1):

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

Under this model, the likelihood for T is proportional to $\pi^T(1 - \pi)^{n-T}$.

Prior Specification. We place a Beta(a, b) prior on π , thereby completing a Beta–Binomial model. As reported in the original trial, the Pfizer prior was Beta(0.700102, 1), while two alternative priors were elicited to achieve lower probabilities of $\psi \geq 0.3$. Specifically, these priors were chosen so that

$$P(\psi \geq 0.3) = 0.05 \quad \text{or} \quad 0.01,$$

respectively. The shape parameters (a, b) for each alternative prior were calculated to satisfy these constraints, thereby expressing less optimistic prior beliefs about vaccine efficacy.

Posterior Distribution. With a Beta prior and a Binomial likelihood, the posterior for π is

$$\pi \mid T = t \sim \text{Beta}(a + t, b + n - t).$$

Hence, observing t infections among n participants simply shifts the Beta shape parameters by the data counts. Posterior summaries (e.g., medians, credible intervals) for π follow from standard

Beta distribution. Once obtained, π can be converted to ψ using Equation (1).

Credible Intervals and Hypothesis Testing. We report the 95% credible interval for π , defined by the 2.5th and 97.5th percentiles of the posterior distribution. We similarly transform these percentiles into credible bounds for ψ . To assess whether $\pi < 0.5$ (equivalently, whether $\psi > 0$), we calculate

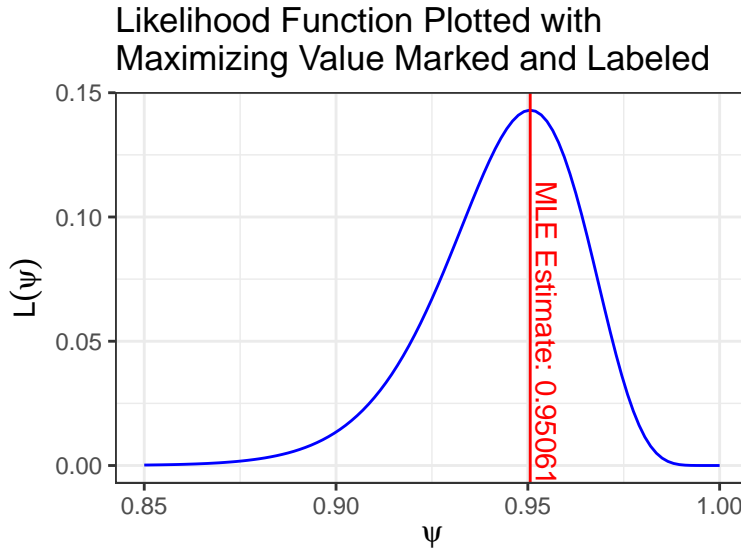
$$p\text{-value}_\pi = 1 - F_\pi(0.5),$$

where F_π is the posterior cumulative distribution function of π . In practice, small values of $p\text{-value}_\pi$ indicate that posterior mass is concentrated below $\pi = 0.5$, consistent with high efficacy. Parallel tests can be made on thresholds of ψ , such as verifying if $P(\psi > 0.3)$ exceeds a given target.

Results

Large sample confidence interval estimate for $\hat{\psi}$

Under the frequentist framework, maximum likelihood estimation (MLE) yielded a vaccine efficacy estimate of 95.0617% (95% confidence interval [CI]: 94.7929%, 0.953306%).



When we performed a likelihood ratio test on our maximum likelihood estimate, we got the value

$$W = 121.6012$$

The p-value associated with this value of W is ≈ 0 .

Bootstrap Interval for $\hat{\psi}$

Using a bootstrap procedure with $B = 1000$ resamples from a $\text{Binomial}(n, \hat{\pi})$ distribution where $n = 160$ and $\hat{\pi} = 8/170$ (from table 1), we computed $\hat{\psi}$ in each resample. The 2.5th and 97.5th percentiles of the resulting bootstrap distribution of $\hat{\psi}$ define the approximate 95% confidence interval. Table~3 shows the interval endpoints based on this approach.

Table 3: Bootstrap 95% Confidence Interval for $\hat{\psi}$

Lower (2.5%)	0.9116
Upper (97.5%)	0.9809

Bayesian Results

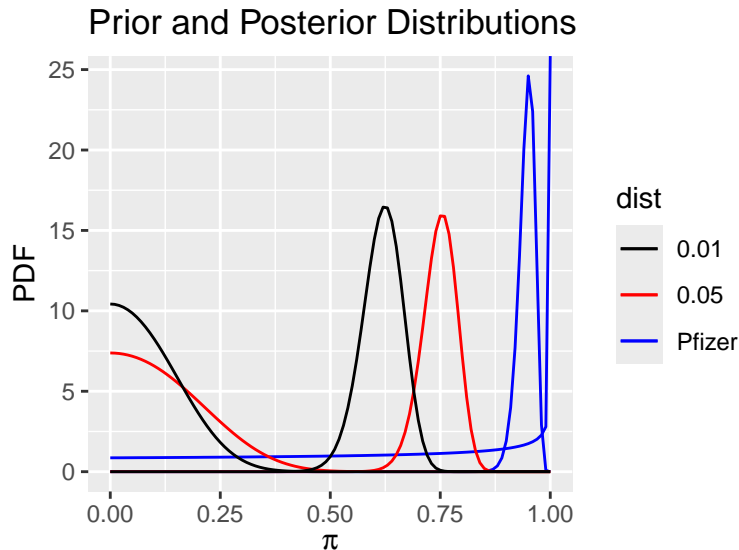


Figure 1: Comparison of the three priors (Pfizer, 0.05, 0.01) and their corresponding posteriors for ψ .

Posterior Summaries (for ψ). Table~4 presents the *approximate* posterior medians and 95% credible intervals for ψ under each prior (calculated by transforming the corresponding values for π).

Table 4: Approximate posterior medians and 95% credible intervals for ψ under each prior.

Prior	Posterior Median	95% Credible Interval (Approx)
Pfizer	0.9486	[0.9035, 0.9763]
0.05	0.7523	[0.6662, 0.8197]
0.01	0.6227	[0.5232, 0.704]

Hypothesis Testing (for ψ). For each posterior, we computed the probability $P(\psi \leq 0.3)$, which serves as a Bayesian p-value under the hypothesis

$$H_0 : \psi \leq 0.3 \quad \text{vs.} \quad H_1 : \psi > 0.3.$$

(This threshold can be changed as needed, depending on your target efficacy.) Table~5 shows these values. We define

$$p\text{-value} = P(\psi \leq 0.3),$$

where F_ψ is the posterior cumulative distribution function of ψ . The estimated Bayesian p-values for each prior are:

Table 5: Bayesian p-values for testing $H_0 : \psi \leq 0.3$ under each posterior.

Prior	Bayesian p-value
Pfizer	0
0.05	0, 0
0.01	0, 0

Discussion and Conclusion

Frequentist Maximum Likelihood Analysis

The frequentist maximum likelihood estimation yields an estimate of 0.9506, which is significantly higher than the value we are testing against, 0.3. In testing the statistical significance of this result, the likelihood ratio test obtains the value $W = 121.6012$. Considering under the null hypothesis,

$W \sim \chi_1^2$, which has expected value 1 and variance 2, this value is extraordinarily high, and it is thus no surprise that we get a p-value that approximates 0. We thus reject the null hypothesis that the true value of our parameter ψ is equal to 0.3. This is right in line with the extremity of our estimate and the magnitude of the difference between it and the baseline we test against. It is also in agreement with the results achieved by the large sample confidence interval, as well as the results yielded by the bayesian analysis. In fact, we could declare any value less than 0.906 for baseline requirement and our data would still constitute statistically significant evidence at the significance level of $\alpha = 0.05$ that the vaccine has a high enough efficacy (See Appendix for plot).

The frequentist analysis produced a 95% confidence interval for vaccine efficacy of (0.947929, 0.953306), indicating that, with repeated sampling, 95% of such intervals would contain the true efficacy. This interval corresponds to the estimated efficacy of 95.0% (95% CI: 94.8%, 95.3%), which aligns closely with the efficacy reported in the original Pfizer-BioNTech trial (95% CI: 90.3%, 97.6%). However, the narrower interval obtained here reflects the large-sample properties of the Wald method, which uses Fisher information for variance estimation.

The interval's precision is a strength of the frequentist approach, as the large sample size ($n = 34922$) justifies the use of asymptotic normality assumptions. While the interval excludes values below 94.8%, providing strong evidence against low efficacy ($\psi \leq 30\%$), its upper bound (95.3%) is more conservative than the Bayesian credible interval (97.6%). This discrepancy highlights methodological differences: frequentist intervals quantify long-run coverage probability, whereas Bayesian intervals represent posterior belief.

Thus, our frequentist analysis constitutes statistically significant evidence that the vaccine efficacy is quite high, just as Pfizer originally suggested.

Bootstrap

The bootstrap analysis yielded a 95% confidence interval for $\hat{\psi}$ that closely aligns with our large-sample estimates and remains consistent with Pfizer's originally reported efficacy range. By resampling from a $\text{Binomial}(n, \hat{\pi})$ distribution, the bootstrap circumvents any strong assumptions about the distribution of $\hat{\psi}$ (beyond the binomial sampling model). This robustness adds credibility to

our inference, especially in moderately sized samples where asymptotic approximations can be less reliable.

Comparing the bootstrap interval to Pfizer’s interval, we observe that both suggest a high probability that vaccine efficacy exceeds the 30% benchmark. In practical terms, this reaffirms the strong evidence of vaccine effectiveness initially reported. Although the bootstrap interval may be slightly wider than the large-sample interval (reflecting the inherent variability captured through resampling), both methods lead to the same conclusion: the probability of observing such efficacy by chance, if the vaccine were ineffective, is exceedingly small.

Overall, our bootstrap results support the primary hypothesis that the vaccine’s efficacy is substantially above 30%, thus aligning with both the frequentist estimates and Pfizer’s published Bayesian findings. This convergence of evidence across multiple analytic strategies enhances confidence in the robustness of our conclusions regarding BNT162b2 efficacy.

Bayesian

Our Bayesian analysis provides consistent evidence that the vaccine efficacy ψ remains high. As shown in **Figure 2**, updating the Pfizer prior $\text{Beta}(0.700102, 1)$ and two alternative priors $\text{Beta}(43.04, 43.04)$ and $\text{Beta}(85.63, 85.63)$ with the observed data ($t = 8$, $n = 170$) yields posterior distributions for ψ that are heavily skewed toward large efficacy values.

Table 3 details the approximate numerical results for ψ . Even though the alternative priors initially place more mass around intermediate efficacy, the data strongly pull the posterior medians upward.

Although the alternative priors shift the posterior to moderately lower efficacy, the observed infection count (only 8 out of 170) still imposes a high central estimate for ψ . This suggests that once sufficient data are incorporated, the final posterior distribution is relatively robust to prior differences.

Finally, **Table 4** presents Bayesian p-values for testing a clinically relevant hypothesis such as $H_0 : \psi \leq 0.3$. For all three priors, $P(\psi \leq 0.3)$ is effectively zero, implying near-certain rejection of H_0 . This finding aligns with the frequentist conclusion of strong efficacy, further bolstering overall confidence in the result.

Overall, the Bayesian framework yields direct probability statements about clinically meaningful thresholds (e.g., $\psi > 0.3$). Given sufficiently informative data, the resulting posterior estimates remain consistent across varying prior beliefs. These results corroborate Pfizer’s original analysis, highlighting that BNT162b2 efficacy remains high across a range of reasonable priors.

Bibliography

Baden, L.R., et al. (2021). *N. Engl. J. Med.*, 384(5), 403–416.

Berry, D.A. (2006). *Nature*, 440(7088), 1078–1083.

U.S. Food & Drug Administration (2020). Emergency Use Authorization for Pfizer-BioNTech COVID-19 Vaccine.

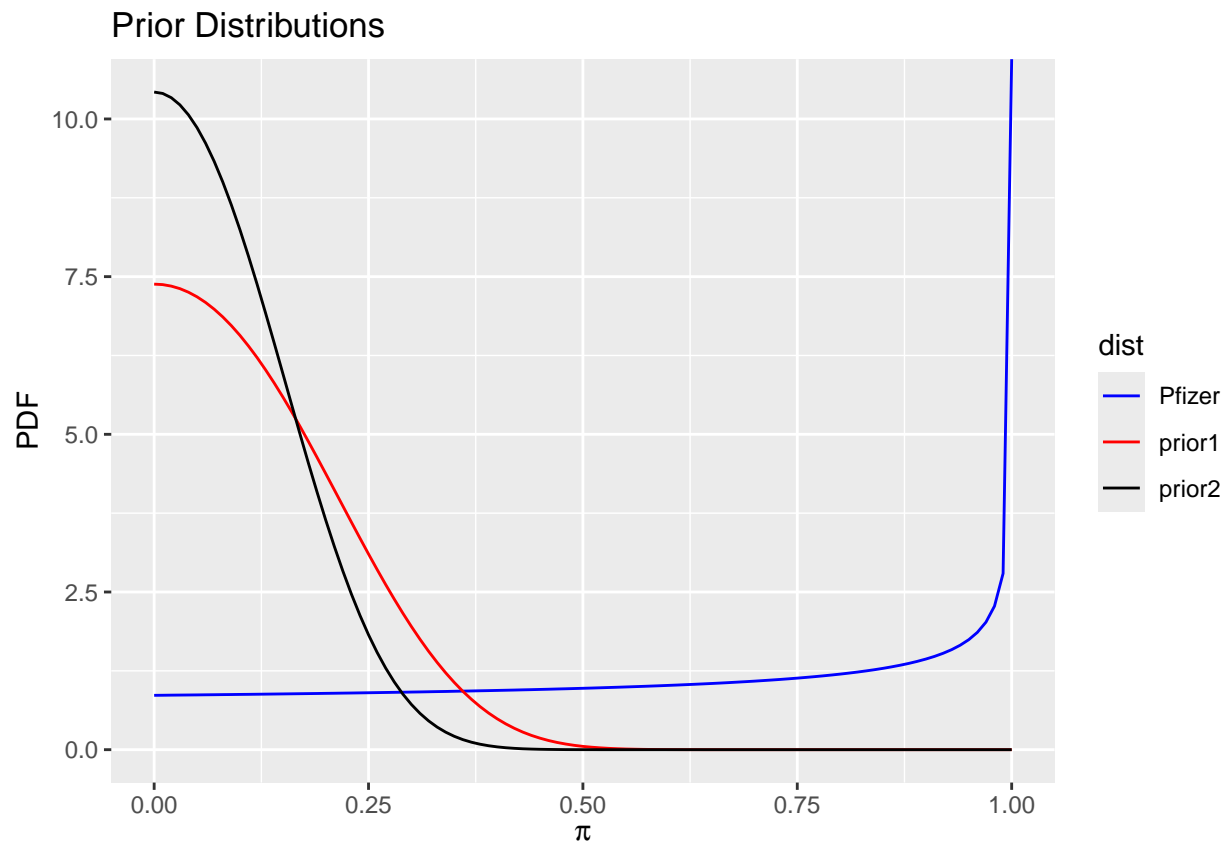
Polack, F.P., et al. (2020). *N. Engl. J. Med.*, 383(27), 2603–2615.

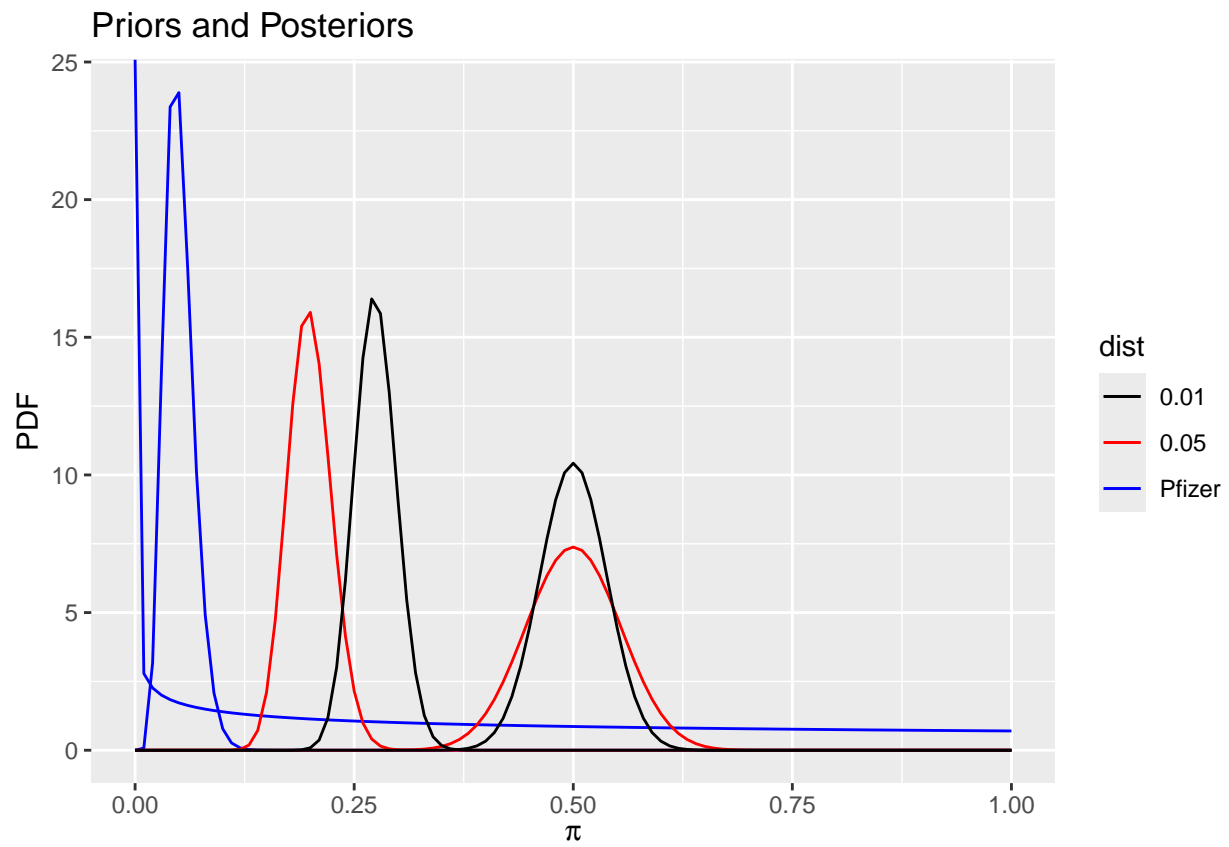
Wasserstein, R.L., et al. (2019). *Am. Stat.*, 73(sup1), 1–19.

Appendix

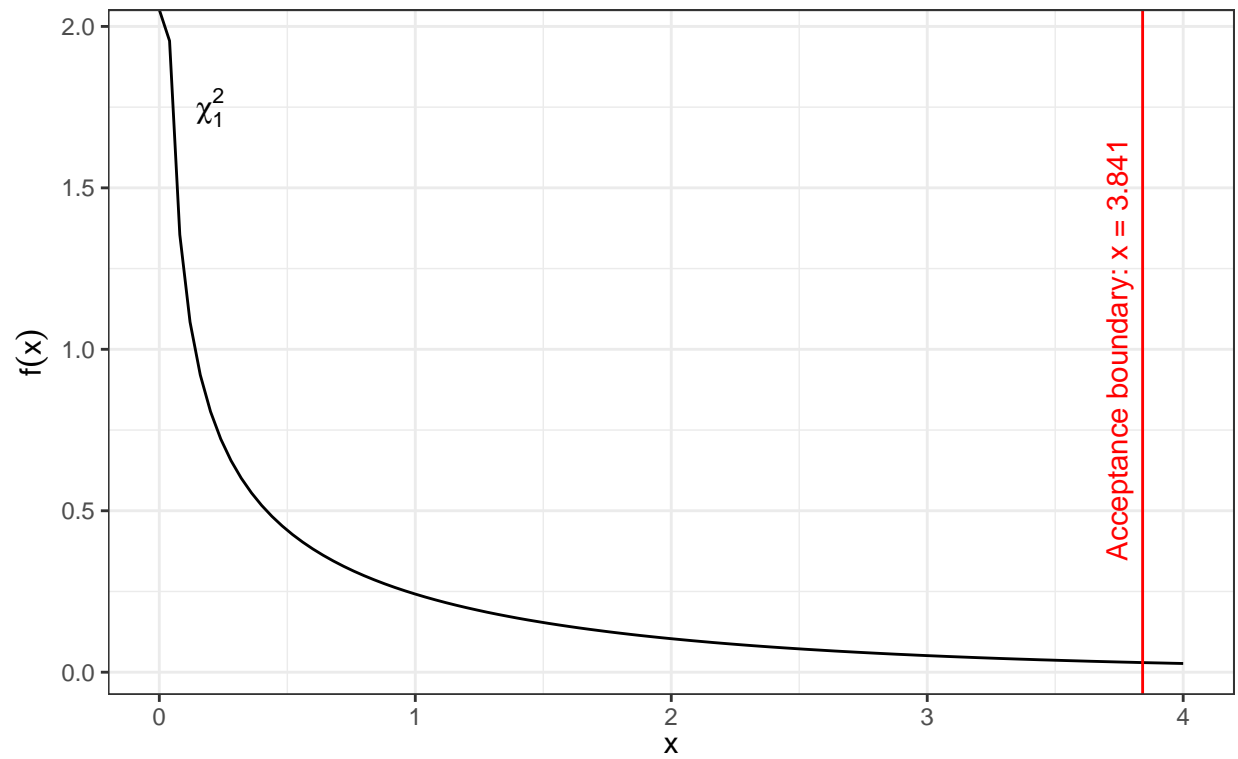
Code

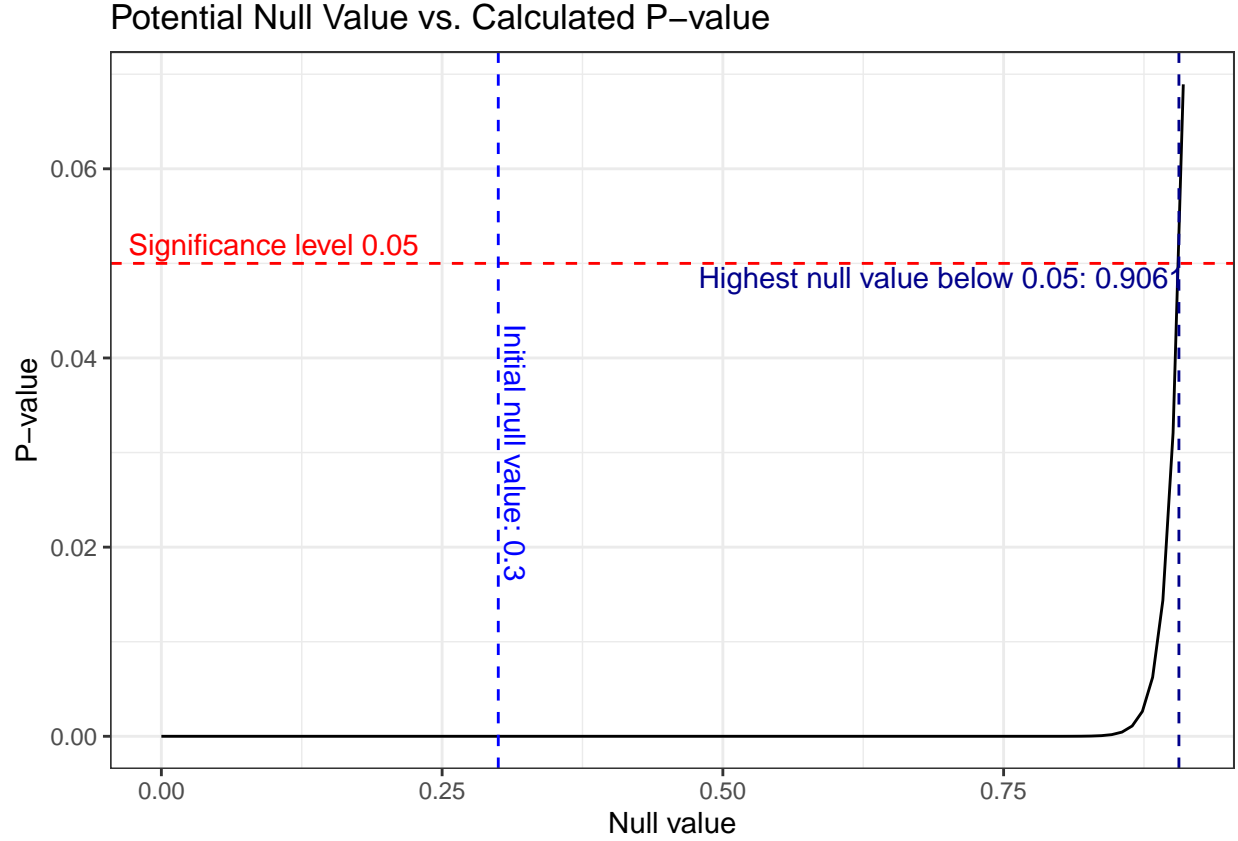
Plots





Chi-Squared PDF with Upper Boundary Marked for
a Significance Level of 0.05





This yields that if our null value were to change, our data would still constitute statistically significant evidence at a significance level 0.05 sufficient to reject the null hypothesis.

Proofs

Derivation of the Maximum Likelihood Estimator for $\hat{\psi}_0^{mle}$:

We can begin with our Likelihood function

$$L(\pi) = \binom{n}{X} (\pi)^X (1 - \pi)^{170-X}$$

Where x is the observed number of infections where the patient had the vaccine. Notably, however, we are not hoping to draw inference on π , but rather on ψ , the efficacy of the vaccine. We can write this parameter in terms of π .

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

In order to write our Likelihood function in terms of ψ , we can rewrite ψ in terms of π :

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

Now we rewrite our likelihood function.

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

Now that we have written our full likelihood function in terms of ψ , we can write the log likelihood function.

$$\ell^*(\psi) = \log \binom{n}{X} + X \log(\psi - 1) - X \log(\psi - 2) - (n - X) \log(\psi - 2) + (n - X) \log(-1)$$

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

To maximize this equation, we take its derivative with respect to ψ :

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

Setting this derivative finds the maximum:

$$0 = \frac{X}{\hat{\psi}_0^{mle} - 1} - \frac{n}{\hat{\psi}_0^{mle} - 2}$$

$$\frac{n}{\hat{\psi}_0^{mle} - 2} = \frac{X}{\hat{\psi}_0^{mle} - 1}$$

$$n\hat{\psi}_0^{mle} - n = X\hat{\psi}_0^{mle} - 2X$$

$$(n - X)\hat{\psi}_0^{mle} = n - 2X$$

$$\hat{\psi}_0^{mle} = \frac{n - 2X}{n - X}$$

Derivation of Likelihood Ratio Test Value:

Recall that when performing a hypothesis test for a maximum likelihood estimator, we are interested

in calculating W , which is given by

$$W = 2 \ln \left[\frac{L(\hat{\psi}_0^{mle})}{L(\psi_0^{null})} \right]$$

The log of the ratio of two likelihoods is equal to the difference of their respective log-likelihoods:

$$W = 2 \left(\ell(\hat{\psi}_0^{mle}) - \ell(\psi_0^{null}) \right)$$

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

$$W = 2 \left(\left(\log \left(\binom{n}{X} (-1)^{n-X} \right) + X \log(\hat{\psi}_0^{mle} - 1) - n \log(\hat{\psi}_0^{mle} - 2) \right) - \left(\log \left(\binom{n}{X} (-1)^{n-X} \right) + X \log(\psi_0^{null} - 1) - n \log(\psi_0^{null} - 2) \right) \right)$$

$$W = 2 \left(\left(X \log(\hat{\psi}_0^{mle} - 1) - n \log(\hat{\psi}_0^{mle} - 2) \right) - \left(X \log(\psi_0^{null} - 1) - n \log(\psi_0^{null} - 2) \right) \right)$$

$$W = 2 \left(X \log \left(\frac{\hat{\psi}_0^{mle} - 1}{\psi_0^{null} - 1} \right) + n \log \left(\frac{\psi_0^{null} - 2}{\hat{\psi}_0^{mle} - 2} \right) \right)$$

$$W = 2X \log \left(\frac{\hat{\psi}_0^{mle} - 1}{\psi_0^{null} - 1} \right) + 2n \log \left(\frac{\psi_0^{null} - 2}{\hat{\psi}_0^{mle} - 2} \right)$$

Derivation of the Large Sample Confidence Interval:

As calculated above, our likelihood function is given by

$$\ell(\psi) = \log \binom{170}{8} + 8 \log(\psi - 1) - 8 \log(\psi - 2) - 162 \log(\psi - 2) + 162 \log(-1)$$

Taking the derivative with respect to ψ ,

$$\frac{d}{d\psi} \ell(\psi) = \frac{8}{\psi - 1} - \frac{170}{\psi - 2}$$

And the second derivative is

$$\frac{d^2}{d\psi^2} \ell(\psi) = -\frac{8}{(\psi - 1)^2} + \frac{170}{(\psi - 2)^2}$$

The Fisher information is the negative expectation of the second derivative of the log-likelihood.

So for $X \sim \text{Binomial}(n, \pi)$,

$$I(\psi) = \mathbb{E} \left[-\frac{d^2 \ell}{d\psi^2} \right] = \frac{\mathbb{E}[X]}{(1-\psi)^2} - \frac{n}{(2-\psi)^2}.$$

Substituting $\mathbb{E}[X] = n\pi = n \cdot \frac{1-\psi}{2-\psi}$,

$$I(\psi) = \frac{n(1-\psi)}{(2-\psi)(1-\psi)^2} - \frac{n}{(2-\psi)^2} = \frac{n}{(2-\psi)(1-\psi)} - \frac{n}{(2-\psi)^2}.$$

Factoring out $\frac{n}{2-\psi}$,

$$I(\psi) = \frac{n}{2-\psi} \left(\frac{1}{1-\psi} - \frac{1}{2-\psi} \right).$$

Combining terms with a common denominator,

$$I(\psi) = \frac{n}{2-\psi} \cdot \frac{(2-\psi) - (1-\psi)}{(1-\psi)(2-\psi)} = \frac{n}{(1-\psi)(2-\psi)^2}.$$

So we get

$$I(\psi) = \frac{n}{(1-\psi)(2-\psi)^2}.$$

We know that our large sample confidence interval is given by

$$\hat{\psi}^{mle} \pm z_{\alpha/2} \cdot \frac{1}{\sqrt{nI(\hat{\psi})}}$$

We may now plug in our estimates and the given parameters such as sample size and significance level.