Pfizer Vaccine Exercise

Just before obtaining emergency approval, in late 2020, Pfizer put out a press release for their investors (!) claiming "a vaccine efficacy rate above 90%" based on "94 confirmed cases of COVID-19 in trial participants". Social media had a field day trying to understand what this meant, also based on previous public documents about the clinical trial. So let's play this game as well. Here is what was known and not known at the time.

- Can assume that there were exactly the same number of patients in the vaccinated and placebo arms. Probably not exactly true, but must be close enough.
- Do not know how many of the $N = 94$ cases were in the two arms (N_V and N_P, $N_V + N_P = N$).
- Vaccine efficiency is defined as $VE = 1 - p_V/p_P$, where p_V (p_P) is the probability of the "event" happening in the vaccinated (placebo) arm.
- Not clear what "above 90%" meant. Maybe a 95% CL?
- Known from previous documents that their analysis was Bayesian on the quantity $\mu = P_V/(P_V + P_P)$, i.e., $VE = (1 - 2\mu)/(1 - \mu)$.
- The prior for μ was a beta distribution $b(0.700102, 1)$.

Assuming the same populations in the two arms, we can rewrite μ as

$$\mu = \frac{N_V}{N_V + N_P} \;=\; \frac{N_V}{N}$$

therefore this is a binomial problem. The binomial probability of observing N_V out of N for a given μ is

$$P(N_V|\mu, N) \;=\; \binom{N}{N_V}\mu^{N_V}(1-\mu)^{N-N_V}$$

The beta distribution (not to be confused with the beta function in math or even the beta function in QFT) is given by

$$b(x; a, b) \;=\; \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$$

and is defined for $0 \le x \le 1$.

Convince yourself that for a beta prior $b(a, b)$, if the observation is N_V, then the posterior for μ is also a beta $b(a + N_V, b + N_P)$[1]. This is an example of a “conjugate prior”, i.e., a prior that yields a posterior of the same functional form. Note that $b(1, 1)$ is the flat prior, and, for what it’s worth, $b(0.5, 0.5)$ is a Jeffrey’s prior. I believe that beta prior are used in binomial problem just as a matter of convenience, i.e., there is no really good reason.
Now let’s play.

- Plot the prior for μ.
- Plot posteriors and calculate 95% confidence limits for a few N_V to try to guess how many vaccinated COVID-19 cases they actually had in their study given the “better than 90%” claim.
- Change the prior to something else reasonable, repeat the previous item, get a feeling for how stable the answer is.
- Repeat again with a frequentist approach and compare.

[1] See for example this.