

That said, we can draw some inferences from the release that may help put this in context. From the press release we know that a total of 94 COVID-19 cases were recorded.

“Upon the conclusion of those discussions, the evaluable case count reached 94 and the DMC performed its first analysis on all cases. “

However, we don't know how many of these come from the control group, and how many come from the treatment group. We also don't know how many total subjects are in the treatment and control arms. We do get two important quotes regarding efficacy.

“Vaccine candidate was found to be more than 90% effective in preventing COVID-19 in participants without evidence of prior SARS-CoV-2 infection in the first interim efficacy analysis

...

The case split between vaccinated individuals and those who received the placebo indicates a vaccine efficacy rate above 90%, at 7 days after the second dose.”

How should we interpret these? Was the observed rate of infection 90% lower in the treatment group, or are we to infer that the true (population parameter) efficacy is at least 90%? I would argue that the wording supports the later. If they were just providing a point estimate why express it as a bound? Why would they phrase it as “indicates a vaccine efficacy rate above 90%” if there was a reasonable probability that the actual vaccine efficacy rate is below 90%?

We can get some additional insight by looking at the [study design](#). It specifies how the interim analysis is to be done. Specifically on pages 102-103, it calls for a Bayesian analysis using a beta binomial model with a weakly-informative prior.

To me, the most compatible statistical translation of their press release is that we are sure with 95% probability that the vaccine's efficacy is greater than 90%. Why “95% probability?” Well, 95% probability intervals are standard for the medical literature if you are doing Bayesian analysis (deal with it), and 95% intervals with 2.5% probabilities on each tail are littered through the design document. They are going to the FDA with these claims, so they will likely stick to the standard evidentiary rules.

Assuming my interpretation is correct, let's back out how many cases were in the treatment group. Conditional on the total number of infections, the number of infections in the treatment group is distributed binomially. We apply the beta prior to this posterior and then transform our inferences from the binomial proportion to vaccine effectiveness. Vaccine effectiveness is one minus the infection rate ratio between the two groups, and the rate ratio is related to the binomial proportion as the odds.

[View Code R](#)

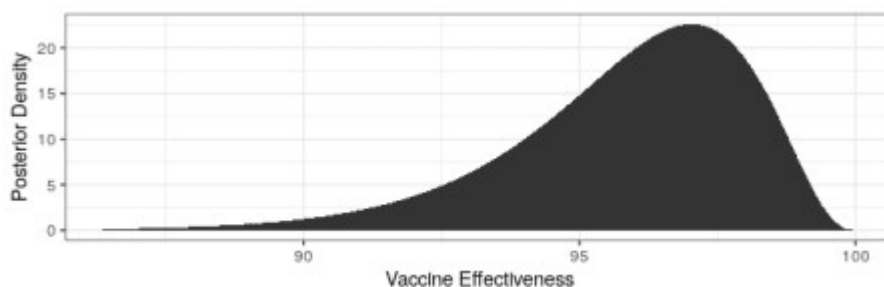
```
> # reference: https://pfe-pfizercom-d8-prod.s3.amazonaws.com/2020-09/C4591001\_Clinical\_Protocol.pdf
>
> # prior interval (matches prior interval on page 103)
> qbeta(c(.025,.975),.700102,1)
[1] 0.005148448 0.964483043
>
>
```

```

> # posterior
> cases_treatment <- 3
> cases_control <- 94 - cases_treatment
> theta_ci <- qbeta(c(.025,.975),cases_treatment+.700102,cases_control+1)
> rate_ratio_ci <- theta_ci / (1-theta_ci)
>
> # effectiveness
> 100 * (1 - rate_ratio_ci)
[1] 98.98688 90.68447
> library(ggplot2)
> xx <- (0:60)/500
> yy <- sapply(xx, function(x) dbeta(x,cases_treatment+.
700102,cases_control+1))
> xx <- 100 * (1 - xx / (1 - xx))
> ggplot() +
+   geom_area(aes(x=xx,y=yy)) +
+   theme_bw() +
+   xlab("Vaccine Effectiveness") +
+   ylab("Posterior Density")

```

The largest number of treatment cases that would have a lower bound greater than 90% is 3, corresponding to 91 cases in the control group. The estimated effectiveness of the vaccine is then 97% with a probability interval from 90.7% to 99.0%. So sure, the effectiveness could be 90% or so, but odds are that it is a **lot** higher as the posterior plot below shows.

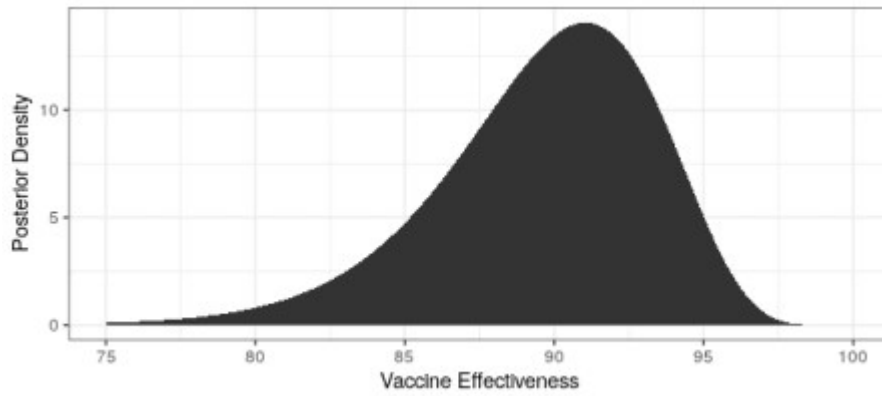


To put this in perspective, consider the rates at which a 97% effective vaccine fails to provide protection, leading to an infection. A 90% effective vaccine has a 3.3 times higher failure rate, so if you vaccinated a population with a 90% effective vaccine and everyone was exposed you'd expect to see 3.3 times more infections compared to if you had used a 97% effective vaccine.

I do note that the analysis plan calls for sequential stopping rules that preserve type I error; however, I don't believe that any reported statistics would be adjusted for that. Unlike frequentist intervals, Bayesian intervals are unchanged no matter how many interim analyses you do.

There is a lot we don't know, and hopefully we will get more scientific clarity in the coming weeks. As it stands now, it seems like this vaccine has efficacy way above my baseline expectations, perhaps even in the 97% range or higher.

I could be wrong in my interpretation of the press release, and they are in fact talking about the sample effectiveness rather than the true effectiveness. In that case, 8 of the 94 cases would have been in the treatment group, and the interval for the true effectiveness would be between 81.6% and 95.6%. The posterior distribution would look pretty darn good, but not quite as nice as the previous one.



It is important to have realistic expectations though. Efficacy is not the only metric that is important in determining how useful the vaccine is. Due to the fact that the study population has only been followed for months, we do not know how long the vaccine provides protection for. There is significant evidence of COVID-19 reinfection, so the expectation is that a vaccine will not provide permanent immunity. If the length of immunity is very short (e.g. 3 months), then it won't be the silver bullet we are looking for. I'd be happy to see a year of immunity and ecstatic if it lasts two.