> "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"

Not being a biostatistician, I was curious how vaccine efficacy is exactly measured. Also, how does the confidence interval look like? Helpfully, Biontech and Pfizer also published a detailed study plan here.

The sample vaccine efficacy can be defined as

$$VE = 1\text{-}IRR = 1 - \frac{c_v/n_v}{c_p/n_p}$$

where n_v and n_p are the number of subjects that got a Covid-19 vaccine and a placebo, respectively, while c_v and c_p are the respective number of subjects that fell ill to the Covid-19 disease. IRR stands for incidence rate ratio and measures the ratio of the share of vaccinated subjects that got Covid-19 to the share in the placebo group.

The press release stated that so far 38955 subjects got the two doses of the vaccine or placebo of which 94 subjects fell ill with Covid-19. Furthermore, the study plan stated that the same number of subjects was assigned to the treatment and control group and let's assume that also in the 38955 subjects analysed so far the ratio is almost equal. An efficacy of at least 90% then implies that from the 94 subjects with Covid-19 at most 8 could have been vaccinated.

The following code computes the IRR and vaccine efficacy assuming that indeed 8 vaccinated subjects got Covid-19.

```
n = 38955 # number of subjects
nv = round(n/2) # vaccinated
np = n-nv # got placebo

# number of covid cases
cv = 8
cp = 94-cv

# percentage of subjects in control group
# who got Covid-19
round(100*cp/np,2)

## [1] 0.44

# percentage of vaccinated subjects
# who got Covid-19
round(100*cv/nv,2)

## [1] 0.04

# incidence rate ratio in % in sample
IRR = (cv/nv)/(cp/np)
round(IRR*100,1)

## [1] 9.3
```

```
# vaccine efficacy in % in sample
VE = 1-IRR
round(VE*100,1)
```

```
## [1] 90.7
```

Assume for the moment this data came from a finished experiment. We could then compute an approximative 95% confidence interval for the vaccine efficacy e.g. using the following formula described in Hightower et. al. 1988

```
arv = cv/nv
arp = cp/np

# CI for IRR
ci.lower = exp(log(IRR) - 1.96 * sqrt((1-arv)/cv + (1-arp)/cp))
ci.upper = exp(log(IRR) + 1.96 * sqrt((1-arv)/cv + (1-arp)/cp))

IRR.ci = c(ci.lower, ci.upper)
round(100*IRR.ci,1)
```

```
## [1]   4.5 19.2
```

```
VE.ci = rev(1-IRR.ci)
round(100*VE.ci,1)
```

```
## [1] 80.8 95.5
```

This means we would be 95% confident that the vaccine reduces the risk of getting Covid-19 between 80.8% and 95.5%. As far as I understood, e.g. the function `ciBinomial` in the package gsDesign allows a more precise computation of the confidence interval:

```
library(gsDesign)
IRR.ci = ciBinomial(cv,cp,nv,np,scale = "RR")
VE.ci = rev(1-IRR.ci)
round(100*VE.ci,1)
```

```
##    upper lower
## 1   81.1  95.4
```

Given that it is only stated that the vaccine is more than 90% effective, the number of Covoid-19 cases may also have been lower than 8 subjects.

The next clean threshold for a press statement would probably be at least 95% effectiveness, which would be exceeded if only 4 vaccinated subjects had Covid-19. So it also seems well reasonable that only 5 vaccinated subjects had Covid-19. This would yield the following vaccine efficacy and confidence interval:

```
cv = 5; cp = 94-cv
VE = 1-(cv/nv)/(cp/np)
round(VE*100,1)
```

```
## [1] 94.4
```

```
IRR.ci = ciBinomial(cv,cp,nv,np,scale = "RR")
```

```
VE.ci = rev(1-IRR.ci)
round(100*VE.ci,1)

##   upper lower
## 1  86.6  97.7
```

Looks even better.

However, those confidence intervals assume a finished, non-adaptive experiment. Yet, the interim evaluations are triggered when the number of Covid-19 cases among the subjects exceeds certain thresholds. The press release states:

> "After discussion with the FDA, the companies recently elected to drop the 32-case interim analysis and conduct the first interim analysis at a minimum of 62 cases. Upon the conclusion of those discussions, the evaluable case count reached 94 and the DMC performed its first analysis on all cases."

> "The trial is continuing to enroll and is expected to continue through the final analysis when a total of 164 confirmed COVID-19 cases have accrued."

I am no expert, but possible the calculation of the confidence interval is not valid for such adaptive rules where the evaluation is triggered by the number of disease cases.

Indeed, Biontech and Pfizer state that they will the assess the precision of the estimated vaccine efficacy using a Bayesian framework with a particular prior distribution described on p. 102-103 of their study plan. Alas, I know very little of Bayesian analysis so I abstain from computing the posterior distribution given the data at hand.

But even absent the full-fledged Bayesian analysis, the numbers really look like very good news.