

# Covid-19 Vaccines

Russell Almond

11/18/2020

Good News! In November, both Pfizer and Moderna announce Phase 3 Vaccine Trials with promising results. With over 30,000 participants in each study they reported the following data.

```
covidVaccines <- tibble(  
  Treatment=c("Placebo", "Vaccine"),  
  Pfizer=c(90,5), ModernaAll=c(95,5),  
  ModernaSevere=c(11,0),  
  N=c(15000,15000))  
covidVaccines.N <- 15000  
kable(covidVaccines)
```

Treatment	Pfizer	ModernaAll	ModernaSevere	N
Placebo	90	95	11	15000
Vaccine	5	5	0	15000

## Measures of Effectiveness

We start with a cross-tab

Treatment	Sick	Healthy	Total
Placebo	SP	HP	NP
Vaccine	SV	HV	NV
Total	NS	NH	N

Odds of getting sick

Placebo: \$ SP/HP \$ Vaccine: \$ SV/HV \$

### Cross Product (Odds) Ratio

$$OR = \frac{SP/HP}{SV/HV}$$

How much does your odds of getting sick increase if you get the placebo instead of the vaccine.

### Risk Ratio

$$RR = \frac{SP/NP}{SV/NV}$$

How much does your probability of getting sick increase if you get the placebo instead of the vaccine.

### Vaccine Effectiveness

$$VE = 100(1 - \frac{1}{RR})$$

## Chi-square test

Null hypothesis is that getting the disease is independent of the vaccine. In other words,  $OR = RR = 1$ .

$$SV/NV = SP/NP$$

Large chi-squared value indicates that cross product rate is not 1 (but doesn't tell if placebo or vaccine is better!

## Z-score test

Another way to work with these data is to calculate probabilities of infection for each group and the standard errors. Then can use the  $z$ -test to compare.

$$p_V = p(S|V) = SV/NV \quad SE(p_V) = \sqrt{p_V(1-p_V)/NV}$$
$$p_P = p(S|P) = SP/NP \quad SE(p_P) = \sqrt{p_P(1-p_P)/NP}$$
$$z = \frac{p_P - p_V}{\sqrt{SE(p_V)^2 + SE(p_P)^2}}$$

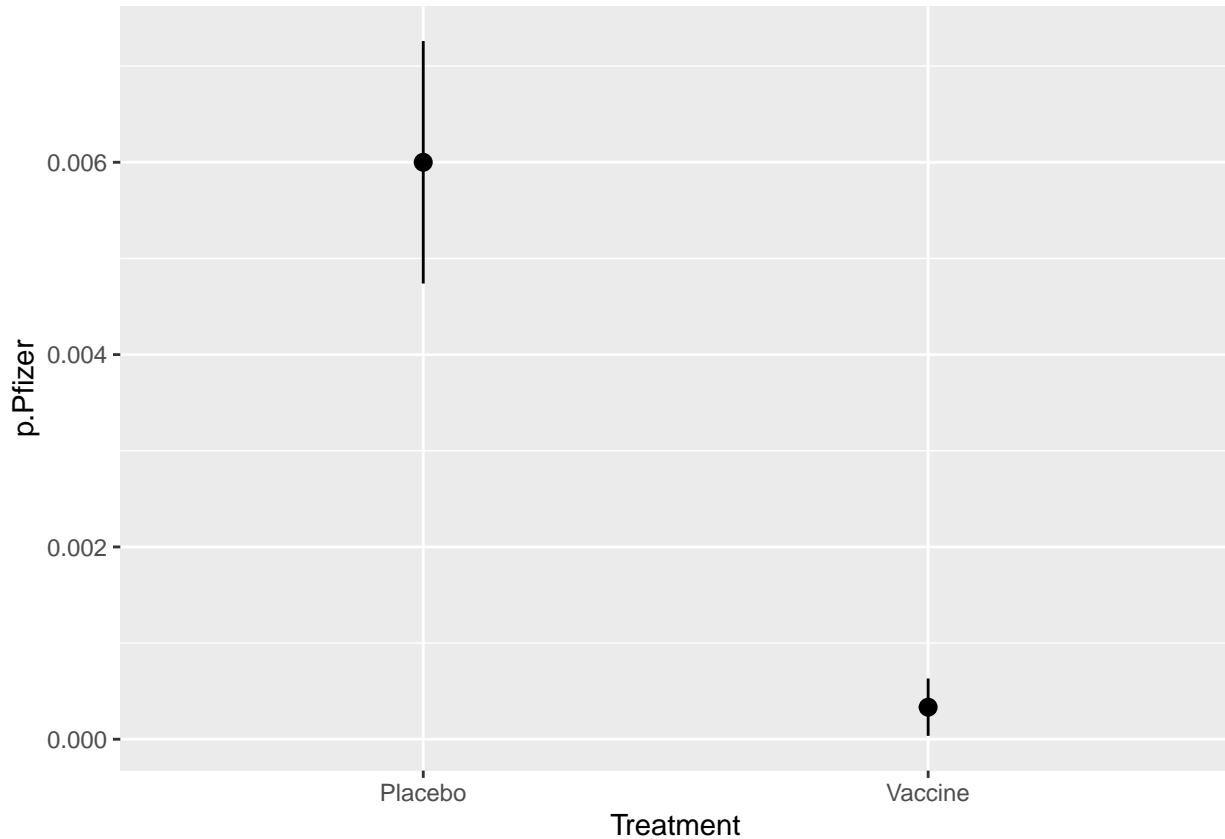
## Pfizer Vaccine

There were around 30,000 volunteers in the Phase 3 trials; 15,000 in each arm.

```
covidVaccines %>%  
  mutate(p.Pfizer=Pfizer/N) %>%  
  mutate(s.Pfizer=sqrt(p.Pfizer*(1-p.Pfizer)/N)) ->  
  covidVaccines  
  
select(covidVaccines, Treatment, contains("Pfizer")) %>% kable(digits=c(4,5))
```

Treatment	Pfizer	p.Pfizer	s.Pfizer
Placebo	90	6e-03	0.00063
Vaccine	5	3e-04	0.00015

```
ggplot(covidVaccines, aes(x=Treatment, y=p.Pfizer, ymin=p.Pfizer-2*s.Pfizer, ymax=p.Pfizer+2*s.Pfizer)) + g
```



### X<sup>2</sup> – Pfizer

In SPSS this is done by producing a cross-tab. We don't have the number of negative cases in each arm of the study, but up to rounding error it is just the sample size, so we will use that.

### Diagnosis \* Vaccine Crosstabulation

		Vaccine				Total	
		Placebo		Active		N	%
Diagnosis		N	%	N	%	N	%
No	No	15000 <sup>a</sup>	99.4%	15000 <sup>b</sup>	100.0%	30000	99.7%
	Yes	90 <sup>a</sup>	0.6%	5 <sup>b</sup>	0.0%	95	0.3%
Total		15090	100.0%	15005	100.0%	30095	100.0%

Each subscript letter denotes a subset of Vaccine categories whose column proportions do not differ significantly from each other at the .05 level.

Figure 1: Pfizer Cross Tab

```
chisq.test(as.matrix(select(covidVaccines,Pfizer,N)))
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
```

## Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	75.813 <sup>a</sup>	1	.000		
Continuity Correction <sup>b</sup>	74.034	1	.000		
Likelihood Ratio	92.281	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	75.811	1	.000		
N of Valid Cases	30095				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 47.37.

b. Computed only for a 2x2 table

Figure 2: Pfizer X2

```
## data: as.matrix(select(covidVaccines, Pfizer, N))
## X-squared = 74.034, df = 1, p-value < 2.2e-16
```

```
p <- pull(covidVaccines,p.Pfizer)
s <- pull(covidVaccines,s.Pfizer)
```

```
z <- (p[1]-p[2])/sqrt(sum(s^2))
pz <- 1-pnorm(z)
cat("Z = ",round(z,2), "p = ",round(pz,3),"\n")
```

```
## Z = 8.75 p = 0
```

```
p <- pull(covidVaccines,p.Pfizer)
RR <- p[1]/p[2]
VE <- 100*(1-1/RR)
cat("Risk Ratio: ",round(RR,2),
    "Vaccine Effectiveness: ",round(VE,1),"\n")
```

```
## Risk Ratio: 18 Vaccine Effectiveness: 94.4
```

## Moderna Vaccine – All Cases

There were around 30,000 volunteers in the Phase 3 trials; 15,000 in each arm.

```
covidVaccines %>%
  mutate(p.ModernaAll=ModernaAll/N) %>%
  mutate(s.ModernaAll=sqrt(p.ModernaAll*(1-p.ModernaAll)/N)) ->
  covidVaccines

select(covidVaccines,Treatment,contains("ModernaAll")) %>% kable(digits=c(4,5))
```

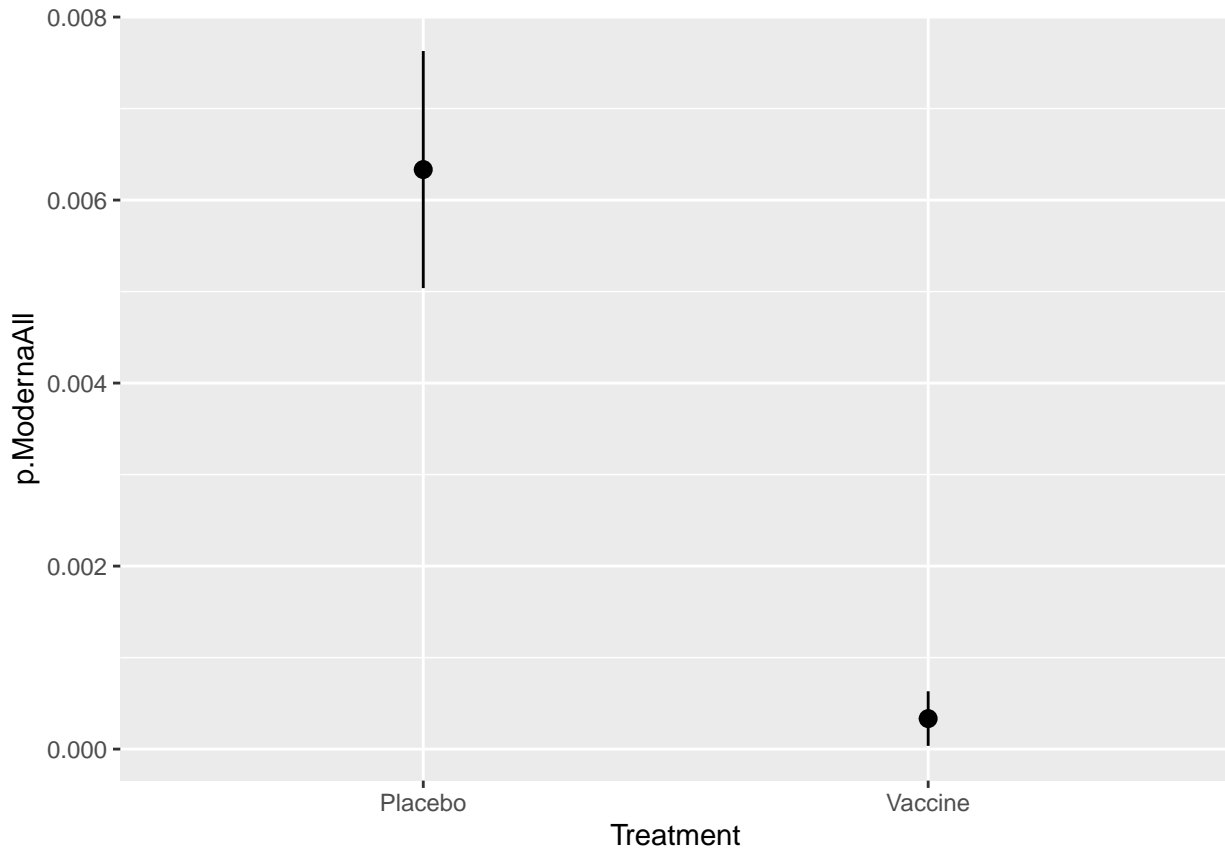
### Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Vaccine (Placebo / Active)	.056	.023	.137
For cohort Diagnosis = No	.994	.993	.996
For cohort Diagnosis = Yes	17.899	7.275	44.035
N of Valid Cases	30095		

Figure 3: Pfizer Risk Ratio

Treatment	ModernaAll	p.ModernaAll	s.ModernaAll
Placebo	95	0.0063	0.00065
Vaccine	5	0.0003	0.00015

```
ggplot(covidVaccines, aes(x=Treatment, y=p.ModernaAll, ymin=p.ModernaAll-2*s.ModernaAll, ymax=p.ModernaAll+
```



### X<sup>2</sup> – Moderna (All Cases)

In SPSS this is done by producing a cross-tab. We don't have the number of negative cases in each arm of the study, but up to rounding error it is just the sample size, so we will use that.

### Diagnosis \* Vaccine Crosstabulation

		Vaccine				Total	
		Placebo		Active		N	%
		N	%	N	%		
Diagnosis	No	15000 <sup>a</sup>	99.4%	15000 <sup>b</sup>	100.0%	30000	99.7%
	Yes	95 <sup>a</sup>	0.6%	5 <sup>b</sup>	0.0%	100	0.3%
Total		15095	100.0%	15005	100.0%	30100	100.0%

Each subscript letter denotes a subset of Vaccine categories whose column proportions do not differ significantly from each other at the .05 level.

Figure 4: Moderna All Cases Cross Tab

```
chisq.test(as.matrix(select(covidVaccines,ModernaAll,N)))
```

```
##
```

```
## Pearson's Chi-squared test with Yates' continuity correction
```

### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	80.732 <sup>a</sup>	1	.000		
Continuity Correction <sup>b</sup>	78.942	1	.000		
Likelihood Ratio	98.657	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	80.729	1	.000		
N of Valid Cases	30100				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 49.85.

b. Computed only for a 2x2 table

Figure 5: Moderna All Cases X2

```
##
## data: as.matrix(select(covidVaccines, ModernaAll, N))
## X-squared = 78.942, df = 1, p-value < 2.2e-16
p <- pull(covidVaccines,p.ModernaAll)
s <- pull(covidVaccines,s.ModernaAll)

z <- (p[1]-p[2])/sqrt(sum(s^2))
pz <- 1-pnorm(z)
cat("Z = ",round(z,2), "p = ",round(pz,3),"\n")

## Z = 9.03 p = 0
p <- pull(covidVaccines,p.ModernaAll)
RR <- p[1]/p[2]
VE <- 100*(1-1/RR)
cat("Risk Ratio: ",round(RR,2),
    "Vaccine Effectiveness: ",round(VE,1),"\n")

## Risk Ratio: 19 Vaccine Effectiveness: 94.7
```

### Moderna Vaccine – Severe Cases

```
covidVaccines %>%
  mutate(p.ModernaSevere=ModernaSevere/N) %>%
  mutate(s.ModernaSevere=sqrt(p.ModernaSevere*(1-p.ModernaSevere)/N)) ->
  covidVaccines

select(covidVaccines,Treatment,contains("ModernaSevere")) %>% kable(digits=c(4,5))
```

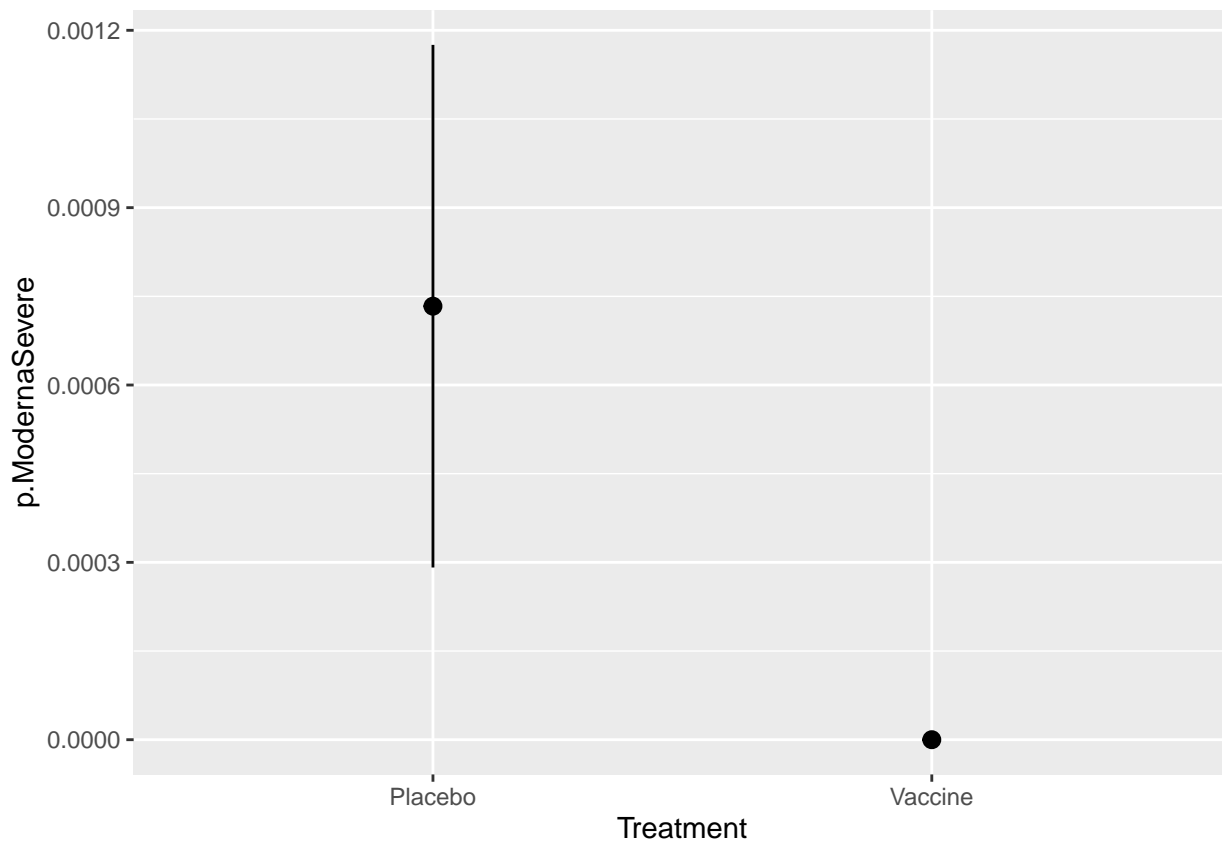
## Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Vaccine (Placebo / Active)	.053	.021	.129
For cohort Diagnosis = No	.994	.993	.995
For cohort Diagnosis = Yes	18.887	7.686	46.408
N of Valid Cases	30100		

Figure 6: Moderna All Cases Risk Ratio

Treatment	ModernaSevere	p.ModernaSevere	s.ModernaSevere
Placebo	11	7e-04	0.00022
Vaccine	0	0e+00	0.00000

`ggplot(covidVaccines, aes(x=Treatment, y=p.ModernaSevere, ymin=p.ModernaSevere-2*s.ModernaSevere, ymax=p.ModernaSevere+2*s.ModernaSevere))`





###  $\chi^2$  – Moderna (All Cases)

In SPSS this is done by producing a cross-tab. We don't have the number of negative cases in each arm of the study, but up to rounding error it is just the sample size, so we will use that.

**Diagnosis \* Vaccine Crosstabulation**

		Vaccine				Total	
		Placebo		Active		N	%
		N	%	N	%		
Diagnosis	No	15000 <sub>a</sub>	99.9%	15000 <sub>b</sub>	100.0%	30000	100.0%
	Yes	11 <sub>a</sub>	0.1%	0 <sub>b</sub>	0.0%	11	0.0%
Total		15011	100.0%	15000	100.0%	30011	100.0%

Each subscript letter denotes a subset of Vaccine categories whose column proportions do not differ significantly from each other at the .05 level.

Figure 7: Moderna Severe Cases Cross Tab

```
chisq.test(as.matrix(select(covidVaccines,ModernaSevere,N)))
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  as.matrix(select(covidVaccines, ModernaSevere, N))
## X-squared = 9.0869, df = 1, p-value = 0.002574
```

```
p <- pull(covidVaccines,p.ModernaSevere)
s <- pull(covidVaccines,s.ModernaSevere)
```

```
z <- (p[1]-p[2])/sqrt(sum(s^2))
pz <- 1-pnorm(z)
cat("Z = ",round(z,2), "p = ",round(pz,3),"\n")
```

```
## Z = 3.32 p = 0
```

```
p <- pull(covidVaccines,p.ModernaSevere)
RR <- p[1]/p[2]
VE <- 100*(1-1/RR)
cat("Risk Ratio: ",round(RR,2),
    "Vaccine Effectiveness: ",round(VE,1),"\n")
```

```
## Risk Ratio: Inf Vaccine Effectiveness: 100
```

Yikes! The estimate for the chances of getting Severe Covid-19 with the virus is 0. Divide by zero error!

But probability zero means impossible. That is not right!

### Continuity Correction

Fix this by adding a continuity correction. We add  $1/2$  to all of the entries in the table.

In particular, this makes the estimated rate for getting severe COVID-19  $\frac{1}{2}/(N + 1)$ .

### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	10.996 <sup>a</sup>	1	.001		
Continuity Correction <sup>b</sup>	9.087	1	.003		
Likelihood Ratio	15.245	1	.000		
Fisher's Exact Test				.001	.000
Linear-by-Linear Association	10.996	1	.001		
N of Valid Cases	30011				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.50.

b. Computed only for a 2x2 table

Figure 8: Moderna Severe Cases X2

```

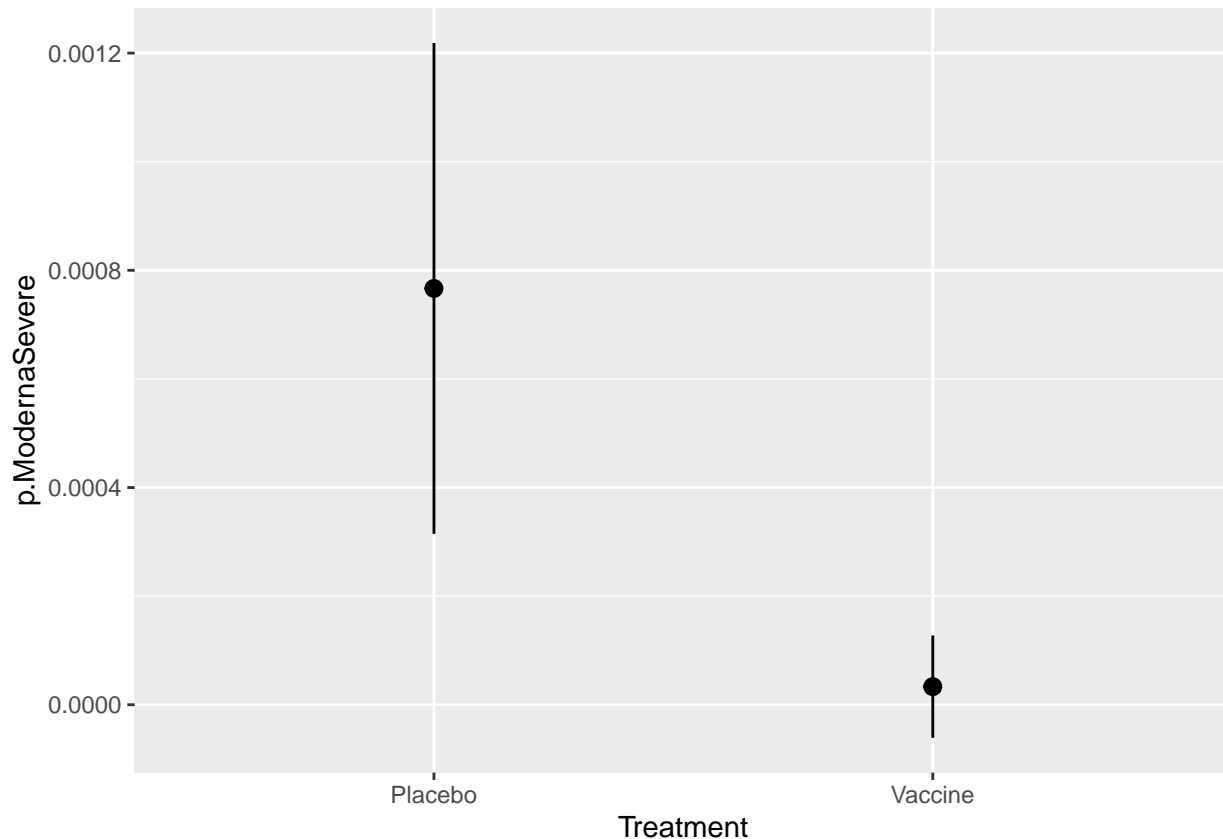
covidVaccines %>%
  mutate(p.ModernaSevere=(ModernaSevere+.5)/(N+1)) %>%
  mutate(s.ModernaSevere=sqrt(p.ModernaSevere*(1-p.ModernaSevere)/(N+1))) ->
  covidVaccines

select(covidVaccines,Treatment,contains("ModernaSevere")) %>% kable(digits=c(5,5))

```

Treatment	ModernaSevere	p.ModernaSevere	s.ModernaSevere
Placebo	11	0.00077	0.00023
Vaccine	0	0.00003	0.00005

```
ggplot(covidVaccines,aes(x=Treatment,y=p.ModernaSevere,ymin=p.ModernaSevere-2*s.ModernaSevere,ymax=p.ModernaSevere+2*s.ModernaSevere))
```



```
p <- pull(covidVaccines,p.ModernaSevere)
RR <- p[1]/p[2]
VE <- 100*(1-1/RR)
cat("Risk Ratio: ",round(RR,2),
    "Vaccine Effectiveness: ",round(VE,1),"\n")
```

## Risk Ratio: 23 Vaccine Effectiveness: 95.7

## References:

- StatNews article on Pfizer vaccine: <https://www.statnews.com/2020/11/09/covid-19-vaccine-from-pfizer-and-biontech-is-strongly-effective-early-data-from-large-trial-indicate/>
- Official Protocol document from Pfizer: <https://www.pfizer.com/science/coronavirus>
- Pfizer Press Release: <https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-announce-vaccine-candidate-against>
- Moderna Press Release: <https://investors.modernatx.com/news-releases/news-release-details/modernas-covid-19-vaccine-candidate-meets-its-primary-efficacy>
- Entries from Andrew Gelman's Blog: <https://statmodeling.stat.columbia.edu/2020/11/16/estimating-efficacy-of-the-vaccine-from-95-true-infections/>

<https://statmodeling.stat.columbia.edu/2020/11/11/the-pfizer-biontech-vaccine-may-be-a-lot-more-effective-than-you-think/>

- *How to use SPSS to obtain Odd Ratio and Relative Risk* <http://brahms.emu.edu.tr/icetin/spss8-RR-OR.pdf>