Cole Adams
ST440 Midterm #1
2/26/2021

Introduction:

Over the last year, COVID-19 has become a global pandemic that has altered the world and changed the paradigm of basic daily life. Developing a vaccine has been an important step in combatting the pandemic, and paving the way for a return to pre-pandemic life. This paper will analyze the trial results of the Moderna vaccine, an mRNA-based vaccine that has been one of the first vaccines brought to market.

Placebo Analysis:

One of the most important indicators for a vaccine is its efficacy, which can be defined as $E = 1 - \frac{\theta_{vaccine}}{\theta_{placebo}}$. , where $\theta_{vaccine}$ is the probability of a randomly-selected person who has received the vaccine contracting the virus. $\theta_{placebo}$ is similar, but the randomly-selected person received the placebo. Our first step will be looking to find both a reasonable likelihood for the data, and a conjugate prior distribution for $\theta_{placebo}$. Given the data states the number of people per group, and then the total number of people per group who contracted the virus, a reasonable likelihood would be binomial. In mathematical terms: $Y|\theta \sim Binomial(N,\theta)$, where N is the total number of people in the placebo group, and θ is the probability of contracting the virus. A reasonable uninformative conjugate prior would be a Beta(1, 1) distribution ($\theta \sim Beta(1,1)$, which is also equivalent to a Uniform(0,1) distribution. This is a good prior in this case because it assigns each θ value the same probability, and only has values possible from 0 to 1. The main assumption we will be making in our model is that the randomly selected patients are independent. This is reasonable since one patient contracting the virus should not impact another, unless they are in contact with each other, and with such a large sample, it is likely this does not play a role. In the next section, we will derive a posterior from our likelihood and prior.

Derivation:

The following steps will show the derivation of the posterior. In these steps, we will use the general prior θ ~ Beta(a,b) in computations.

$$Y \mid \theta \sim Binomial(N, \theta)$$
$$\theta \sim Beta(\alpha,\ \beta)$$

$$P(\theta \mid Y) \ \propto \ \left[\theta^{Y}(1-\theta)^{N-Y}\right] * \theta^{\alpha-1}(1-\ \theta)^{\beta-1}$$
$$P(\theta \mid Y) \ \propto \ \theta^{Y+\alpha-1}(1-\ \theta)^{N-Y+\beta-1}$$
$$P(\theta \mid Y) \ \propto \ Beta(\alpha+Y, \beta+N-Y)$$
$$With\ \theta \sim Beta(1,\ 1):$$
$$P(\theta \mid Y) \ \propto \ Beta(Y+1, N-Y+1)$$

The posterior distribution of $\theta_{placebo}$ of all patients can be visualized below:

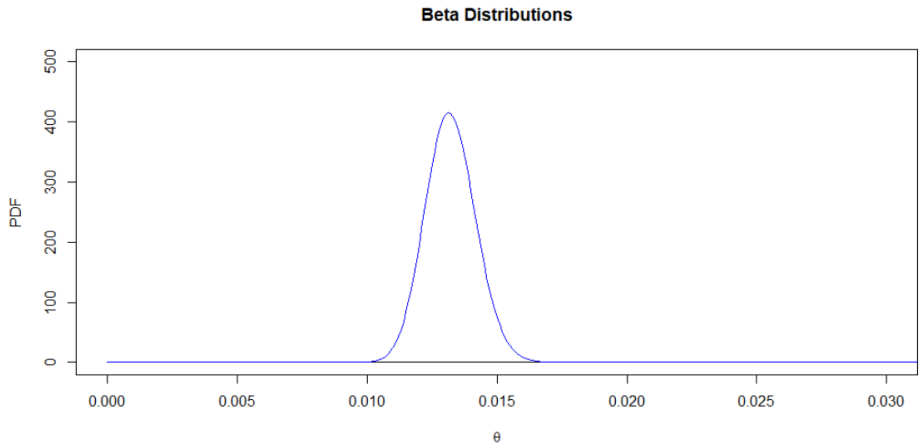

Figure 1: The $\theta_{placebo}$ distribution for all patients with prior Beta(1,1).

Efficacy Analysis:

First, it is important to note that the $\theta_{vaccine}$ had an identical likelihood and prior as $\theta_{placebo}$, with the only difference being that the group was the vaccine group and not the placebo group. Using Monte Carlo sampling, with samples of 100,000 points, a Bayesian analysis of the efficacy of the vaccine was performed. This was done with both the original prior θ ~ Beta(1,1), as well as with four other priors to determine the sensitivity of the results to the prior. The results, specifically the mean and 95% credible set are tabulated and outputted below. As can be seen, there is slight variation in the results depending on the prior, but not massive

differences, with the largest differences being between .03-.04. Thus, we can conclude that the results are mildly sensitive to the prior selected.

	Mean	2.5%	97.5%
Beta(1,1)	0.935	0.892	0.967
Beta(2,2)	0.930	0.886	0.964
Beta(5,5)	0.916	0.867	0.953
Beta(5,2)	0.916	0.867	0.953
Beta(2,5)	0.930	0.886	0.964

Figure 2: Bayesian analysis of vaccine efficacy, with different priors

.

Our efficacy results using our pre-selected prior had a mean of .935, with a credible set (.892, .967). Thus, we can say that we are 95% confident that the efficacy of the vaccine falls within that set, given the prior. Now, we will visualize the five distributions of the Monte Carlo samples. It can be seen that a couple priors have close resulting distributions, but that there are slight differences throughout the collective group.

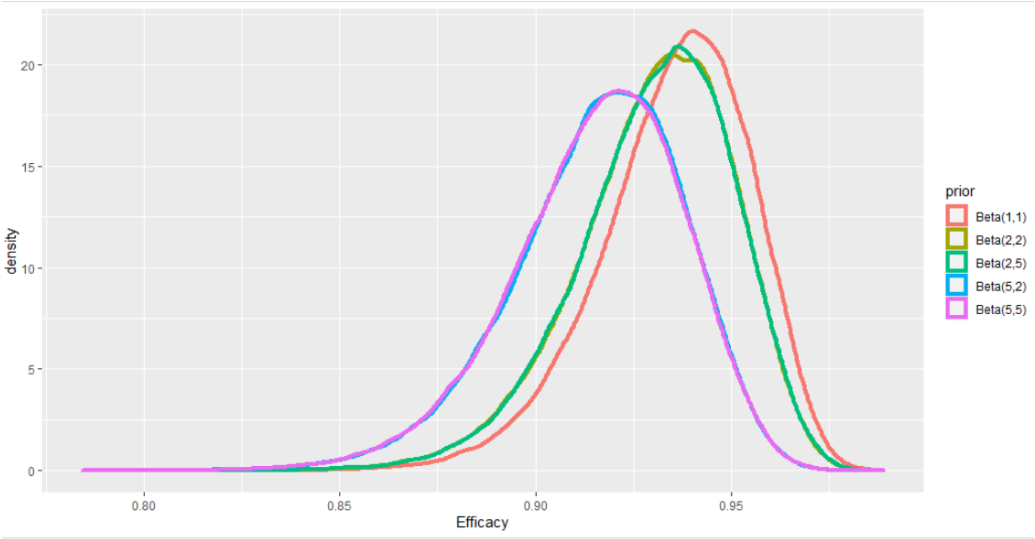

Figure 3: Visualizations of the five monte carlo sampling result distributions

Hypothesis Test:

Our goal is to test whether the efficacy was greater than .70. To do this, the mean of the Monte Carlo sample distribution above .70 will be calculated, and if this value is .95 or higher, we can conclude statistically that the vaccine has at least a .70 efficacy. This value, when

calculated, is ~1. This lines up with Figure 3, as in that plot, there appears to be very little of the distribution at .8, let alone .7. A value of ~1 can lead us to conclude that we are nearly 100% confident that the vaccine's efficacy is above .70. This result is not sensitive to our prior, as all five versions hold the same value of ~1, as can be seen in Figure 3 as well.

Subgroup Analysis:

Next, analysis will be done on the two different subgroups: patients that are white and patients that come from communities of color. Similar to what was done for all patients, the same posterior distribution was computed with modified N and θ values being inputted to match each group's statistics. Then Monte Carlo sampling was done to build a sampling distribution. This visualization and analysis are shown below in Figure 4 and Figure 5.

	Mean	2.5%	97.5%
White	0.924	0.8705187	0.9630342
Communities of Color	0.951	0.8582289	0.9943021

Figure 4: The Bayesian analysis of vaccine efficacy, split by group

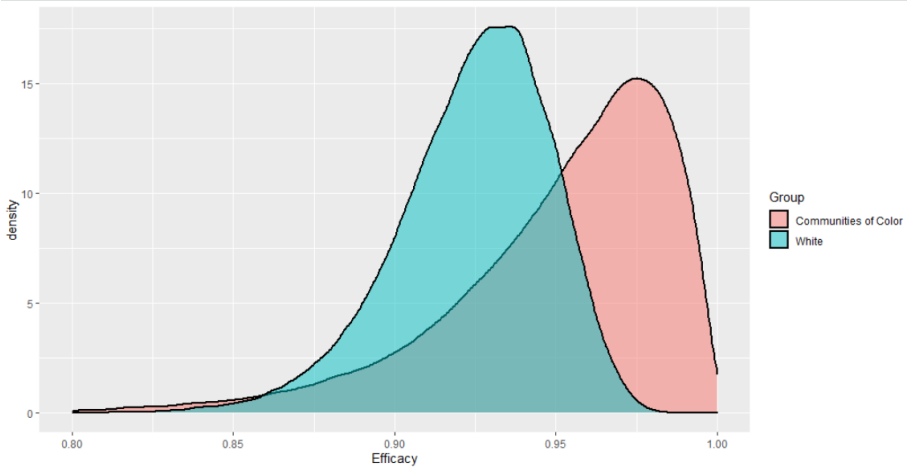

Figure 5: Visualization of sampling distribution for vaccine efficacy, split by group

It can be seen that the vaccine had a higher efficacy on communities of color by nearly .03. To test if this difference is significant, we will compute the mean of the communities of color Monte Carlo sample that is greater than the Monte Carlo sample of white patients. This value is .77047, and since this is not greater than .95, we cannot say that the vaccine's efficacy on communities of color is statistically greater than white patients.

Code:

```r
# Cole Adams Midterm Exam
# ST440
set.seed(2)

#Problem Set-Up
pplaceboAll <- 14073
pplaceboWhite <- 8916
pplaceboCol <- 5132

iplaceboAll <- 185
iplaceboWhite <- 144
iplaceboCol <- 41

pvaccineAll <- 14134
pvaccineWhite <- 9023
pvaccineCol <- 5088

ivaccineAll <- 11
ivaccineWhite <- 10
ivaccineCol <- 1

#Efficacy Check for Later

ETotal <- 1 - ((ivaccineAll / pvaccineAll) / (iplaceboAll / pplaceboAll))
EWhite <- 1 - ((ivaccineWhite / pvaccineWhite) / (iplaceboWhite / pplaceboWhite))
ECol <-1 - ((ivaccineCol / pvaccineCol) / (iplaceboCol / pplaceboCol))

#Distribution for Priors/ Posteriors
theta <- seq(0,1,.0001)

#Alpha / Betas for Prior
a1<-1
b1<-1

a2 <- 2
b2 <- 2
a3 <- 5

b3 <- 5

a4 <- 5
b4 <- 2

a5 <- 2
b5 <- 5

#Beta Prior
betaPrior<- dbeta(theta, a1, b1)



#Posterior for Vaccine and Placebo
posteriorPlaceboAll <- dbeta(theta,a1 + iplaceboAll,pplaceboAll - iplaceboAll + b1)
posteriorVaccineAll <- dbeta(theta,a1 + ivaccineAll,pvaccineAll - ivaccineAll + b1)

#Plotting (use for #2)
plot(NA, type = "l", xlim = c(0,.03), ylim = c(0, 500), xlab=expression(theta), ylab = "PDF", main = "Beta Distributions")
lines(theta,betaPrior)
lines(theta, posteriorPlaceboAll, col = "Blue")


#############
```

```r

#3)
MC <- 100000
ppMC <- rbeta(MC,a1 + iplaceboAll,pplaceboAll - iplaceboAll + b1)
pvMC <- rbeta(MC,a1 + ivaccineAll,pvaccineAll - ivaccineAll + b1)

ppMC2 <- rbeta(MC,a2 + iplaceboAll,pplaceboAll - iplaceboAll + b2)
pvMC2 <- rbeta(MC,a2 + ivaccineAll,pvaccineAll - ivaccineAll + b2)

ppMC3 <- rbeta(MC,a3 + iplaceboAll,pplaceboAll - iplaceboAll + b3)
pvMC3 <- rbeta(MC,a3 + ivaccineAll,pvaccineAll - ivaccineAll + b3)

ppMC4 <- rbeta(MC,a4 + iplaceboAll,pplaceboAll - iplaceboAll + b4)
pvMC4 <- rbeta(MC,a4 + ivaccineAll,pvaccineAll - ivaccineAll + b4)

ppMC5 <- rbeta(MC,a5 + iplaceboAll,pplaceboAll - iplaceboAll + b5)
pvMC5 <- rbeta(MC,a5 + ivaccineAll,pvaccineAll - ivaccineAll + b5)

eff1 <- (1 - pvMC / ppMC)
eff2 <- (1 - pvMC2 / ppMC2)
eff3 <- (1 - pvMC3 / ppMC3)
eff4 <- (1 - pvMC4 / ppMC4)
eff5 <- (1 - pvMC5 / ppMC5)

meansAll <- rbind(round(mean(eff1),3), round(mean(eff2),3), round(mean(eff3),3), round(mean(eff4),3), round(mean(eff5),3))
rownames(meansAll) <- c("Beta(1,1)", "Beta(2,2)", "Beta(5,5)", "Beta(5,2)", "Beta(2,5)")
colnames(meansAll) <- "Mean"
quantileAll <- rbind(round(quantile(eff1,c(0.025,0.975)),3), round(quantile(eff2,c(0.025,0.975)),3), round(quantile(eff3,c(0.025,0.975)),3), round(q

finalAll <- cbind(meansAll, quantileAll)

finalAll %>%
  kbl() %>%
  kable_classic(full_width = F, html_font = "Cambria")

#3 Plots:
df <- data.frame(prior=factor(rep(c("Beta(1,1)", "Beta(2,2)", "Beta(5,5)", "Beta(5,2)", "Beta(2,5)"), each=100000)), Efficacy = c(eff1,eff2,eff3,eff
ggplot(df, aes(x = Efficacy, color = prior)) + geom_density(alpha = 0.5, size = 1.5)


############

#4)
mean(eff1 > .7)
mean(eff2 > .7)
mean(eff3 > .7)
mean(eff4 > .7)
mean(eff5 > .7)
```

```r
#############

#4)
mean(eff1 > .7)
mean(eff2 > .7)
mean(eff3 > .7)
mean(eff4 > .7)
mean(eff5 > .7)

#5)
ppMCWhite <- rbeta(MC,a1 + iplaceboWhite,pplaceboWhite - iplaceboWhite + b1)
pvMCWhite <- rbeta(MC,a1 + ivaccineWhite,pvaccineWhite - ivaccineWhite + b1)

effWhite <- (1 - pvMCWhite / ppMCWhite)

meanWhite <- round(mean(effWhite),3)
quantWhite <- quantile(effWhite,c(0.025,0.975),3)

ppMCCol <- rbeta(MC,a1 + iplaceboCol,pplaceboCol - iplaceboCol + b1)
pvMCCol <- rbeta(MC,a1 + ivaccineCol,pvaccineCol - ivaccineCol + b1)

effCol <- (1 - pvMCCol / ppMCCol)

meanCol <- round(mean(effCol),3)
quantCol <- quantile(effCol,c(0.025,0.975),3)

meanDiff <- rbind(meanWhite, meanCol)
quantDiff <- rbind(quantWhite, quantCol)

finalDiff <- cbind(rbind(meanWhite, meanCol), rbind(quantWhite, quantCol))
colnames(finalDiff)[1] <- "Mean"
rownames(finalDiff) <- c("White", "Communities of Color")


finalDiff %>%
  kbl() %>%
  kable_classic(full_width = F, html_font = "Cambria")

mean(effCol > effWhite)

df2 <- data.frame(Group=factor(rep(c("White", "Communities of Color"), each=100000)), Efficacy = c(effWhite,effCol))
ggplot(df2, aes(x = Efficacy, fill = Group)) + geom_density(alpha = 0.5, size = 1) + xlim(.8,1)
```