

Most observers seem surprised that the regimen with only half an initial dosage showed a substantially larger efficacy. Some theories for this result are sketched in [this Nature news article](#). An obvious question is: How statistically robust is the 90% efficacy reported for the smaller dosing regimen?

This post first performs several educated guesses to reverse-engineer the underlying case numbers from the press release. Then we follow [Biontech/Pfizer's Bayesian analysis](#) approach to compare the posterior distributions of AstraZeneca's two dosage regimens with those of the Biontech/Pfizer and Moderna trials.

Let s_1 denote the share of the 131 Covid cases that accrued in the first dosage regimen. If we ignore rounding errors in the stated efficacy, we can compute it from the equation that determines the average efficacy of 70% of both dosage regimens by solving

$$0.9 * s_1 + 0.62 * (1 - s_1) = 0.7$$

which yields $s_1 = 2/7 = 28.6\%$.

So while approximately 28.6% of the 131 Covid-19 cases come from the first, smaller dosing regimen only $2741 / 11637 = 23.6\%$ of the participants are from that regimen. Given that the smaller dosing regimen has higher efficacy, I would rather have suspected that its share of the Covid-19 cases is smaller than the 23.6% share of the participants.

The result means that the share of participants in the control group that got Covid-19 is larger in the first dosing regimen than in the second one.

Looking at AstraZeneca's press release in more detail, we read

The pooled analysis included data from the COV002 Phase II/III trial in the UK and COV003 Phase III trial in Brazil.

A description of the UK and Brazil trials reveals that the UK trial had both dosing regimens, while the Brazil trial only had the second larger dosing regimen. A quick internet search did not confirm that the Covid-19 risk was smaller in Brazil than in the UK. Yet, the UK trial may well have started earlier than the Brazil trial, which would give participants more time to catch Covid-19. That might explain the relative high Covid-19 case proportion in the smaller dosing regimen.

For the moment, we will ignore integer constraints and thus compute that $m_1 = (2/7) * 131 = 37.42$ cases were from the smaller dosing regimen. As a next step, we want to compute the number of Covid-19 cases mv_1 from the vaccinated treatment group of the smaller dosing regimen, using the reported efficacy of $VE_1=90\%$.

We first compute for the smaller dosing regimen the helper parameter θ_1 that shall measure the share of the m_1 Covid-19 cases that were from vaccinated subjects. To compute it, we need to make an assumption about the subject split between treatment and control group. Since no information is given in the press release, let us *assume* that AstraZeneca has a 1:1 assignment to treatment and control group, like Biontech/Pfizer. Then we have (see e.g. [here](#))

$$\theta_1 = (1 - VE_1) / (2 - VE_1)$$

From this we can compute mv_1 , as well as the cases from the control group mc_1 in the smaller dosing regimen. The following R code computes in this fashion all relevant case numbers for both dosing regimens:

```

m = 131

m1 = m*(2/7)
VE1 = 0.9
theta1 = (1-VE1)/(2-VE1)

mv1 = theta1*m1
mc1 = m1-mv1

m2 = m*(5/7)
VE2 = 0.62
theta2 = (1-VE2)/(2-VE2)

mv2 = theta2*m2
mc2 = m2-mv2

rbind(
  smaller_dosing = c(m=m1, mv=mv1,mc=mc1, VE=VE1),
  larger_dosing = c(m=m2, mv=mv2,mc=mc2, VE=VE2)
)

##               m           mv           mc    VE
## smaller_dosing 37.42857   3.402597 34.02597 0.90
## larger_dosing  93.57143 25.766046 67.80538 0.62

```

OK, the obvious problem remains that these case counts are no integer numbers. That is probably due to the fact that the reported efficacy percentages are rounded. So we have to guess integer numbers that yield efficacy values that are plausible to have been rounded to the reported numbers. As a guess, let us just round the numbers above to the nearest integers:

```

# Integer guess: just round case numbers above
m1 = 37; mv1 = 3; mc1 = 34
m2 = 94; mv2 = 26; mc2 = 68

# Compute resulting efficacies
c(
  VE1 = 1-(mv1/mc1),
  VE2 = 1-(mv2/mc2),
  VE  = 1-(mv1+mv2)/(mc1+mc2)
)

##           VE1           VE2           VE
## 0.9117647 0.6176471 0.7156863

```

This means these assumed case counts would yield 91.2% efficacy in the small dosing regimen, 61.8% efficacy in the large dosing regimen and a 71.6% average efficacy. Seems roughly consistent with the stated numbers. Of course, a lot of guesses went into this computation.

So if we assume 3 of 37 Covid-19 cases in the small dosage regimen were vaccinated compared to 26 of 94 in the large dosage regimen, is the difference in these proportions significant? While I am no expert on non-parametric tests (economists tend to mostly run regressions), the R function [prob.test](#) seems on first sight appropriate. So let's just run it:

```
prop.test(x=c(mv1,mv2),n=c(m1,m2))

##
##      2-sample test for equality of proportions with continuity
correction
##
## data:  c(mv1, mv2) out of c(m1, m2)
## X-squared = 4.8083, df = 1, p-value = 0.02832
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.34049235 -0.05053698
## sample estimates:
##      prop 1      prop 2
## 0.08108108 0.27659574
```

The p-value of 2.8% indeed gives some support for the hypothesis that the true efficacy is larger in AstraZeneca's small dosing regimen. (Of course, just taking this p-value at face value might be slightly p-hackish.)

As a final step, let us compare AstraZeneca's results with those reported by Biontech/Pfizer and Moderna using the Bayesian approach suggested by Biontech/Pfizer's study plan (see [my previous post](#) for details). [Biontech reported](#) that out of 170 Covid-19 cases 8 subjects were vaccinated and [Moderna's press release](#) states 5 out of 95 cases.

The code below shows for each trial / dosing regimen the posterior distribution for the vaccine efficacy using the prior Beta distribution specified by Biontech/Pfizer and our guessed numbers for AstraZeneca.

```
# Helper function
theta.to.VE = function(theta) (1-2*theta)/(1-theta)

# Parameters of Biontech/Pfizer's prior distribution
a0 = 0.700102; b0 = 1
grid = tibble(
  study=c("Biontech/Pfizer","Moderna","AstraZeneca-1","AstraZeneca-2"),
  m=c(170,95,37,94),
  mv = c(8,5,3,26),
  mc=m-mv) %>%
  tidyr::expand_grid(theta = seq(0,0.5,by=0.002)) %>%
  mutate(
    VE = theta.to.VE(theta),
    density = dbeta(theta, shape1 = a0+mv, shape2=b0+mc)
  )

# Show all 4 posterior distributions
ggplot(filter(grid, VE > 0.2), aes(x=VE, y=density, fill=study,
color=study)) +
  geom_area(alpha=0.5,position = "identity")+
  ggtitle("Estimated posterior efficacy of different vaccines / dosing
regimens")
```



Hmm, for me subjectively the posterior for AstraZeneca's small dosing regimen looks actually better than I would have suspected after reading the press release. Of course, a lot of guesses went into those curves.

Overall, I personally would consider AstraZeneca's preliminary results good news. In particular, if one also accounts for the substantially lower expected price and the following statements from the press release:

no hospitalisations or severe cases of the disease were reported in participants receiving the vaccine.

The vaccine can be stored, transported and handled at normal refrigerated conditions (2-8 degrees Celsius/ 36-46 degrees Fahrenheit) for at least six months.