On the Effects of Misclassification in Estimating Efficacy With Application to Recent COVID-19 Vaccine Trials.

John P. Buonaccorsi
Dept. of Mathematics and Statistics
University of Massachusetts-Amherst

ABSTRACT Understandably, the recent trials for COVID-19 vaccines have garnered a considerable amount of attention and (as of this writing) vaccinations are about to begin. The popular summaries give infection rates in the vaccinated and placebo and estimated efficacy, which for the two trials we focus on (Moderna and Pfizer) are both near 95%. This paper explores the potential effects of possible false positives or false negatives (misclassification) in the COVID-19 diagnosis with specific application to the Moderna and Pfizer trials. The general conclusion, fortunately, is that these potential misclassifications almost always would lead to underestimation of the efficacy and that correcting for false positives or negatives will lead to even higher estimated efficacy.

1 Introduction

There has been a flurry of recent results coming from large trials of potential COVID-19 vaccines. Two of them are the Moderna and Pfizer trials, both very large studies with volunteers randomized in roughly equal numbers to the vaccine or a placebo. The main object of this note is to explore the potential effects of misclassification of COVID-19 status on the estimated efficacy of the vaccines and assess correction for hypothesized probabilities of false positives and false negatives. This will be done using a combination of analytical results and numerical explorations. Some of the results here parallel some of those in De Smedt et al. (2018) who examined the same question in assessing vaccine effectiveness. As they point out, however, efficiency is not the same as efficacy so there are some differences in the approach. The focus here is specifically on efficacy and in particular on scenarios based on the rates involved in the recent COVID-19 trials.

Since the two aforementioned trials form the basis of values used in our numerical investigations, we summarize them first. At the time of this writing the reported values for the two trials were as shown in Table 1, where

N_v = number of subjects in the vaccinated group

N_p = number of subjects in the placebo group

I_v = number of subjects infected in vaccinated group

I_P = number of subjects infected in placebo group

$R_v = I_v/N_v$ = infection rate in vaccinated group

$R_p = I_p/N_P$ = infection rate in placebo group

Study	N_v	N_p	I_v	I_p	R_v	R_p	$\widehat{Eff}$
Moderna	15000	1500	5	90	.00033	.006	94.44%
Pfizer	17411	17511	8	162	.00046	.0093	95.04%

Table 1: Basic results from two COVID-19 vaccine trials

The efficacy of the vaccine is estimated using

$$\widehat{Eff} = 100 * [1 - (R_v/R_p)]$$

The sample sizes in the Moderna trial are approximate based on the reported number of approximately 30000 total participants, with a goal to randomize equal numbers to the placebo and vaccine groups. Although these are not exact, with these

large sample sizes, moderate changes will have little effect on the results and exact values are not critical to our main objective.

From the statistical perspective the goal is to estimate the true efficacy of the vaccine, defined as be

$$Eff = 100(1 - (\pi_v/\pi_P)) = 100(1 - \rho), \tag{1}$$

where $\rho = \pi_v/\pi_p$,
π_v = probability a random person from the population gets COVID-19 if vaccinated, and
π_p = probability a random person from the population gets COVID-19 if given the placebo.

NOTE: There is no assumption here that the probability a vaccinated person gets COVID-19 is the same across people. That is not realistic. Each person has a different exposure leading to a different probability and even two people with the exact same exposure could have different probability of getting the disease. If you vaccinated everyone then the probability a random person selected from the population gets the disease is π_v which is the average of the individual probabilities over the population. When people are randomly selected and then randomized to the placebo or vaccinated the probability of a vaccinated person (which is over both random selection and assignment) is π_v. And given the large samples involved the resulting infection count I_v can be treated as Binomial with sample size N_v and probability π_v Similar comments apply for the placebo group.

Assuming, to start, that the presence or absence of COVID-19 can be determined without error, then R_v is unbiased for π_v and R_p is unbiased for π_p. Since we are dealing with a ratio this does not necessarily mean that R_v/R_p is unbiased for $\rho = \pi_v/\pi_R$ but the biases in this situation will be small, estimated to be essentially zero in both trials; see the appendix.

A standard error (SE) and confidence interval (CI) for ρ, and subsequently for the efficacy, which were absent in the general reporting of the trial results can be obtained using methods for estimating a ratio; see Buonaccorsi[2001]. The confidence intervals given below use the generally more dependable Fieller's method but these are almost identical to those from the delta method, which use the estimated efficacy $\pm 1.96SE$. One can also take a Bayesian approach to the problem. Based on posted documents associated with the approval process this leads to a credible region of [90.3,97.6] for the efficacy of the Pfizer vaccine, close to the Fieller interval below.

Trial	SE	CI
Moderna	2.50%	[89.1,99.3]
Pfizer	1.8%	[91.3,98.5]

It is clear here that that the estimated efficacies are highly precise in both trials.

2 Bias due to diagnostic error in the COVID-19 test

Since there are concerns about the fallibility of COVID-19 diagnoses, and the potential for false negatives in particular, the natural question is what effect potential misdiagnoses of the disease might have on the estimated efficacy; a topic that was absent from the initial press on these trials. In the discussion below we will draw on the summary of the role of misclassification in estimating a proportion/prevalence given by Buonaccorsi(2010,Ch.2). Additional references can be found therein.

2.1 Non-differential misclassification

To, start assume the sensitivity and specificity are the same in the vaccinated and the placebo group (what is called non-differential misclassification), with

$$\text{PFP} = \text{P(false positive) and PFN} = \text{P(false negative)}.$$

The sensitivity of the test is 1 - PFN and the specificity is 1 - PFP.

The probability of observing a positive result for the given diagnostic method is
P(positive result|positive)P(positive) + P(positive result|negative)P(negative),
where | denotes "given". For the vaccinated group this leads to

$$p_v = (1 - PFN)\pi_v + PFP(1 - \pi_v) = \pi_v(1 - PFN - PFP) + PFP,$$

while for the placebo group

$$p_p = (1 - PFN)\pi_p + PFP(1 - \pi_p) = \pi_p(1 - PFN - PFP) + PFP.$$

Note that p will equal the true π if PFN and PFP are both 0.

The observed positive rates, R_v and R_p are estimating p_v and p_p respectively, rather than the corresponding π_v and π_p and the observed sample ratio R_v/R_p is estimating essentially

$$\rho_{obs} = \frac{p_v}{p_p} = \frac{\pi_v(1 - PFN - PFP) + PFP}{\pi_p(1 - PFN - PFP) + PFP}. \tag{2}$$

Once again, given that we are dealing with a ratio, this is an approximate/limiting result, not exact. However, biases based on it are very accurate given the sample sizes

involved. In the later numerical illustrations this was verified by simulation (results not shown here). Note that if one had small sample sizes some modification may be needed but that is not the case for the situations under consideration here.

Defining $c = 1 - PFP - PFN$ for convenience, the bias in the ratio R_v/R_p as an estimator of $\rho = \pi_v/\pi_p$ is approximately

$$B_R = \frac{p_v}{p_p} - \frac{\pi_v}{\pi_p} = \frac{PFP(\pi_p - \pi_v)}{\pi_p(\pi_p c + PFP)} = \frac{(1 - (\pi_v/\pi_P))PFP}{\pi_p * c + PFP}.$$

The bias in the estimated efficacy is

$$Bias_{eff} = 100 * (1 - \rho_{obs}) - 100 * (1 - \rho) = -100 * (\rho_{obs} - \rho) = -100 B_R.$$
$$= \frac{-Eff * PFP}{\pi_p * (1 - PFP - PFN) + PFP}.$$

Two main conclusions emerge:

- If the probability of a false positive is 0 then this bias is 0.
- The efficacy will almost always be underestimated, The only case where this would not happen is if the denominator is negative. This cannot happen if 1-PFP - PFN is greater than 0 which would certainly be true for any sensible test procedue.

To explore the issue further, we evaluated the bias numerically over a number of different settings. Based on the results from the trials we fixed the infection rate in the placebo group, π_p, to be .006 to start. We then varied the efficacy over a variety of values between 50% and 95% (which in turn determines the vaccine rate π_v) and used a grid of values for PFP and PFN. Since it is generally accepted that the sensitivity is high and hence the PFP is small, the PFP ranged from 0 to .01 by .001 while PFN went from 0 to .1 in steps of .01. For each combination we evaluated $100(1-(p_v/p_p))$, which is what is being estimated using the observed rates. The process was then repeated by doubling π_p to .012 to reflect higher incidence rates (which would occur over longer observation times) but with the same efficacy.

Figures 1 - 3 display results corresponding to an efficacy of 95%, 70% and 50%, respectively. In each case we plot $100(1 - (p_v/p_p))$ versus PFP where each indicates a different value of PFN. The black lines are for the original setting with $\pi_p = .006$ and the red lines for the cases with $\pi_v = .012$. The true efficacy is indicated by the solid horizontal line (e.g., at 95 when efficacy = 95%).

A few conclusions emerge immediately:

1. The efficacy can quickly be grossly underestimated as PFP grows. The rapidity which with this happens is more marked at the smaller rates ($\pi_p = .006$ with a corresponding π_v depending on the efficacy.) In practice, only the smallest PFPs matter.

2. For a given PFP the bias is rather insensitive to the probability of a false negative.

3. The bias is greatest at smaller incidence rates (black lines).

Since it is the probability of false negatives that is the main concern with COVID-19 diagnoses and the PFP is typically quite small, these results are reassuring. This may or may not be the case for other vaccine trials depending on the diagnostic measure used and the prevalence rates involved.

2.2 Adjusting for misclassification

We can also examine the issue for the two trials by exploring directly what an adjusted estimate of efficacy is for given misclassification probabilities.

For given PFP and PFN, corrected estimates of the infection rates (see for example equation (2.3) in Buonaccorsi(2010)) are given by

$$\hat{\pi}_v = \frac{R_v - PFP}{1 - PFP - PFN}$$

and

$$\hat{\pi}_p = \frac{R_p - PFP}{1 - PFP - PFN}$$

for the vaccinated and placebo groups respectively.

From these a corrected estimate of the efficacy, denoted $\widehat{Eff}_c$ is given by

$$\widehat{Eff}_c = 100 * (1 - \frac{\hat{\pi}_v}{\hat{\pi}_p}) = 100 * (1 - \frac{R_v - PFP}{R_p - PFP}).$$

It can also be expressed as

$$\widehat{Eff}_c = 100 * \frac{R_p - R_v}{Rp - PFP} = \widehat{Eff}\frac{R_p}{R_p - PFP},$$

where, recall, $\widehat{Eff}$ is the original estimate of efficacy based on the observed rates.

- *Although the false negative rate is needed to get a corrected estimate of disease prevalence in each group, it does not enter into getting a corrected estimate of efficacy!* This is similar to what was found in De Smedt et al. (2018) for estimating effectiveness. This only applies with non-differential misclassification.

- Note that if $PFP > R_v$ then the corrected estimate of π_v is negative and similarly the estimate of π_p is negative if $PFP > R_p$. Obviously, the underlying probabilities we are trying to estimate cannot be negative. So, with very small estimates of disease rates (as is the case in both the trials we are using for illustration) obtaining a corrected estimate will be problematic if the PFP becomes too large. This would occur (i.e, both estimates would be negative) if PFP is greater .006 in the Moderna trial or greater than .008 in the Pfizer trial.

We applied the correction to the original data for varying levels of PFP (but chosen to keep the corrected estimated rates positive) for each of the trials as shown in the black lines in Figures 4 and 5. The red lines in those figures come from doubling the infection rates in the original data to see to what extent the very small rates in the original data influenced the conclusions. Recall the probability of a false negative does not enter into the correction.

The story is pretty clear here and was expected based on the earlier theoretical bias discussion. Correcting for misclassification due to non-zero probability of a false positive always leads to a larger estimated efficacy. For these two trials even a modest PFP leads to estimates approaching 100%.

2.3 Differential Misclassification

2.3.1 Biases

Differential misclassification occurs if the probabilities of false positives and false negatives differ in the placebo and vaccinated groups. In this case the analytical assessment of bias induced by misclassification becomes more complicated. We now denote the probabilites involved by:
PFP_v = P(false positive) in vaccinated group
PFN_v = P(false negative) in vaccinated group
PFP_p = P(false positive) in placebo group, and
PFN_p = P(false negative) in placebo group.

The estimator based on the oberved infection rates, $\widehat{Eff}$, is now is estimating

$$100 * (1 - \frac{\pi_v(1 - PFN_v - PFP_v) + PFP_v}{\pi_p(1 - PFN_p - PFP_p) + PFP_p})$$

rather than the true efficacy. This leads to a bias of

$$B = 100 * (\frac{\pi_v}{\pi_p} - \frac{\pi_v(1 - PFN_v - PFP_v) + PFP_v}{\pi_p(1 - PFN_p - PFP_p) + PFP_p}).$$

Unlike the case with non-differential misclassification there are no easy expressions leading to easy insight into the nature of the bias. However it is easy to explore numerically. Similar to what was done earlier, the infection rate in the placebo group was set to .006 with efficacies of 95% (π_v = .0003),70% (π_v = .0018) and 50% (π_v = .006). We looked at all combinations resulting from varying PFP_v from 0 to .01 by .001, PFP_p from 0 to .01 by .001, PFN_v from 0 to .10 by .02 and PFN_p from 0 to .10 by .02.

Figures 6 - 8 show the nature of the bias as a function of PFP_v for specified values of PFP_p not equal to PFP_v. The plots show what is being estimated using the observed rates compared to the true efficacy (indicated by the solid black horizontal line) where at each value of PFP_v the multiple values correspond to all the combinations of the two probabilities of false negatives, some of which are equal, others which are not.

- The nature of the bias is determined primarily by the probabilities of false positives involved.
- The general nature of the bias is not that sensitive to the probabilities of false negatives, although, interestingly, the spread of biases of the different probabilities of false negatives is greater the smaller the probability of a false positive is in the placebo group.
- At the higher efficacies (95% and 70%) the efficacy is almost always underestimated. The exception is in a few unrealistic (and uninteresting) settings where the PFP in the vaccinated group is 0 or very small and probability of a false positive in the placebo groups becomes larger than it. But even there the overestimation is slight.
- At an efficacy of 50% (the vaccine does not help, but does not harm) there can be more substantial overestimate of the efficacy in certain situations. As above, this happens when the probability of a false positive in the is vaccinated group is low and the probability of a false positive in the placebo group is higher. Then the overestimation could be substantial. However, this is a a very unlikely scenario and we note that at the smaller PFP of .001 in the placebo group, the efficacy will be underestimated. So, in the most realistic scenarios we still see underestimation.

2.3.2 Correcting for misclassification

The general strategy of correcting for misclassification using information on the false positive and false negative rates is similar to the non-differential cases. The corrected estimates of the infection rates in the two groups are

$$\hat{\pi}_v = \frac{R_v - PFP_v}{1 - PFP_v - PFN_v}$$

for the vaccinated group and

$$\hat{\pi}_p = \frac{R_p - PFP_p}{1 - PFP_p - PFN_p}$$

for the placebo group.

From these the corrected estimate of the efficacy, denoted $\widehat{Eff}_c$ is given by

$$\widehat{Eff}_c = 100 * (1 - \frac{\hat{\pi}_v}{\hat{\pi}_p}).$$

Similar to treating bias, there are no simple expressions showing the nature of the correction and unlike the case of non-differential misclassification we do need information about both the false positive and false negative rates, for each group. This was also noted by De Smedt et al. (2018) for estimating effectiveness.

Figures 9 and 10 show the effects of correcting for misclassification using the Moderna and Pfizer data using all combination of probabilities of false positives ranging from 0 to .0003 in each group and the probabilities of false negatives in the two groups ranged from 0 to .01. Once again, we see that the correction almost always moves the original estimated efficacy closer to 100%. The only exception to this were very small drops in the estimated efficacy in the unrealistic scenarios where the PFP in the vaccinated group was 0 but was positive in the placebo group.

3 Discussion

The discussion here has focused on the impact of potential misclassification in the COVID-19 diagnosis on the estimated efficacy and how one can correct the estimated efficacy using information about misclassification rates. The conclusion is that in these settings the misclassification will almost always lead to underestimation of the efficacy and correcting for misclassification would almost always push the estimate even higher than the approximately 95% observed in both trials.

A fuller treatment of the problem would also address the issue of how to get standard errors and confidence intervals for the efficacy using either known or estimated probabilities of false positives and false negatives. This is relatively standard using inferences for ratios. In the case of using estimated misclassification rates, the particulars of the analysis will depend on whether the validation is internal or external. This is tangental to our main focus here and will be addressed in a separate paper.

Appendix

It is well known that the expected value of af a ratio is not exactly the ratio of the expected value. So, in our setting with no misclassification, the expected value of R_v/R_p is not exactly π_v/π_v. But, using an approximation (see for example Mood, Graybill and Boes (1974,p.181)) in this situation, where R_p and R_v are uncorrelated, the bias in the ratio is approximately $\pi_v Var(R_p)/\pi_p^3$. Using the data this can be estimated by $p_v(p_p(1-p_p)/N_v)/p_p^3$, which will be small for large N_v. The estimated bias in the efficacy is -100*(bias in the ratio). This is estimated to be -.062 and -.03 for the Moderna and Pfizer trials respectively, which for all practical purposes is 0 compared to the estimated efficacies near 95.

References

Buonaccorsi, J. (2001). "Fieller's Theorem". *Encycolopedia of Environmetrics,* .

Buonaccorsi, J. P. (2010) *Measurement Error: Models, Methods and Applications.* Chapman & Hall.

De Smedt, D, Merell, E, Macina, D., Perez-Villar,S., Andrews,N. and K. Bollaerts (2018). "Bias due to differential and non-differential disease and exposure misclassification in studies of vaccine effectiveness",
PLoS ONE 13(6): e0199180. https://doi.org/10.1371/journal.pone.0199180.

Mood, A., Graybill, F. and D. Boes.(1974). *Introduction to the Theory of Statisitics.* McGraw-Hill.

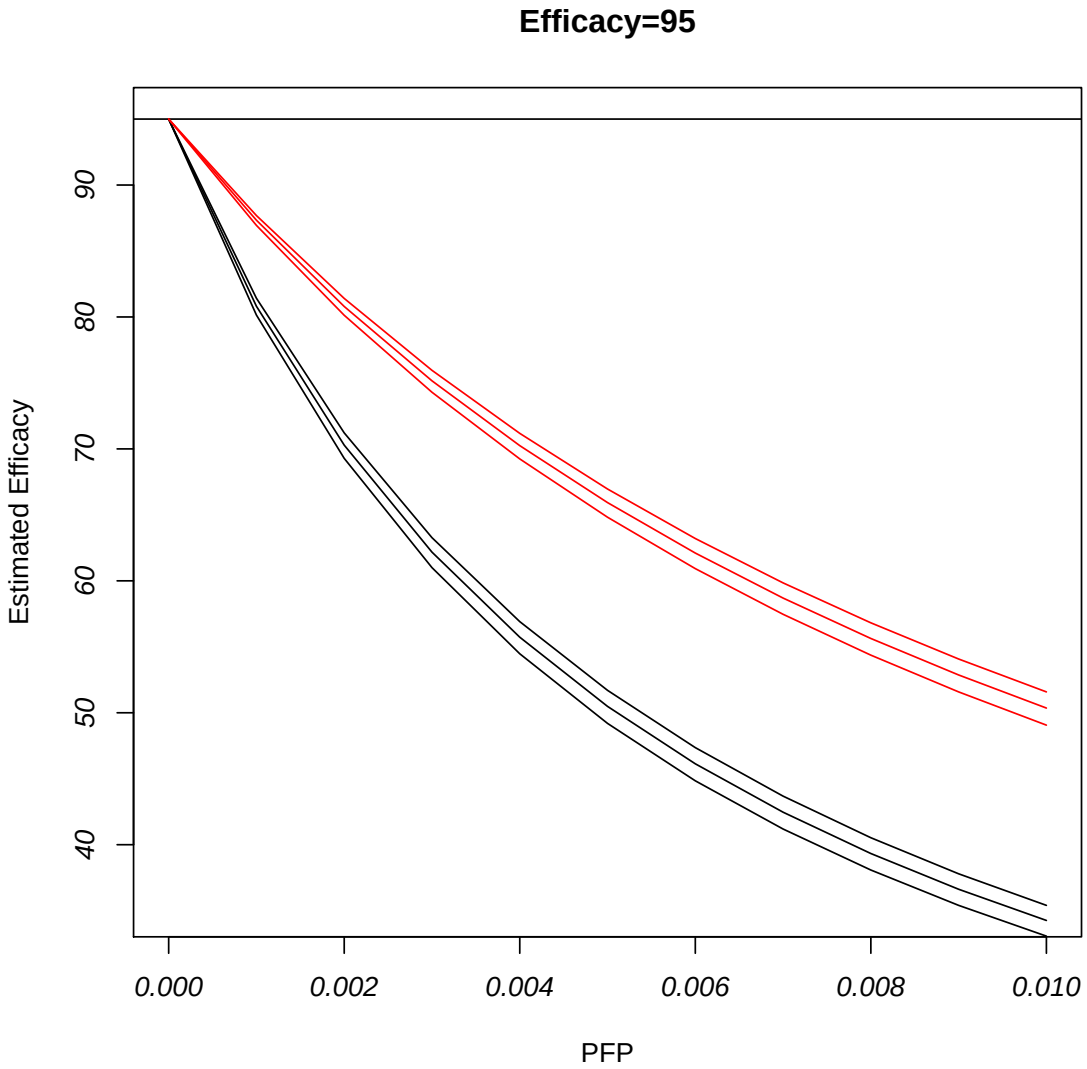

Figure 1: Illustration of bias due to misclassification at efficacy = 95% (solid horizontal line). Plot of what is estimated using the observed rates as a function of PFP. Black lines are for different PFNS based on $\pi_p = .006$ with $\pi_v = .0003$. Red lines are based on $\pi_p = .012$ with $\pi_v = .0006$.

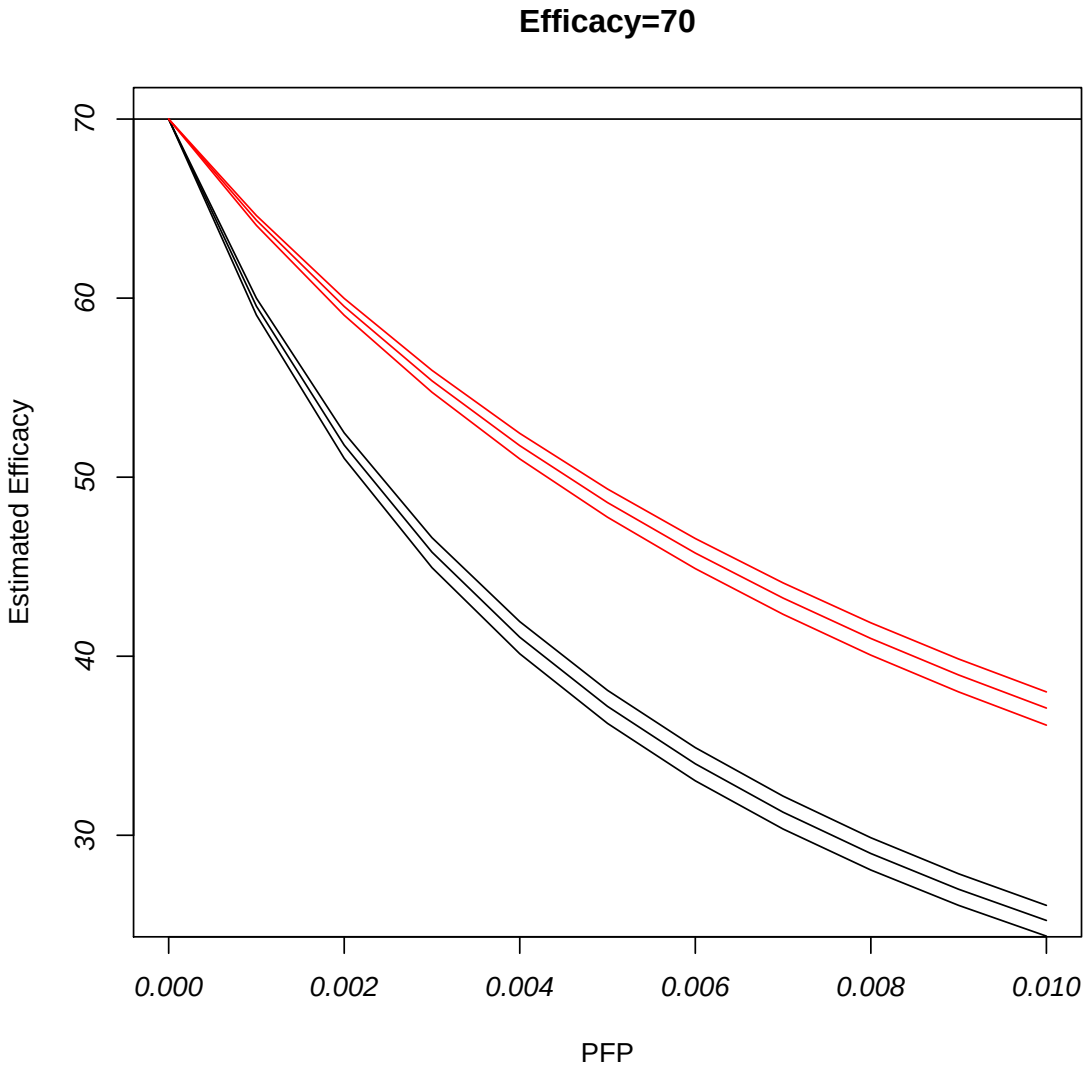

Figure 2: Illustration of bias due to misclassification at efficacy = 70% (solid horizontal line). Plot of what is estimated using the observed rates as a function of PFP. Black lines are for different PFNS based on $\pi_p = .006$ with $\pi_v = .0003$. Red lines are based on $\pi_p = .012$ with $\pi_v = .0006$.

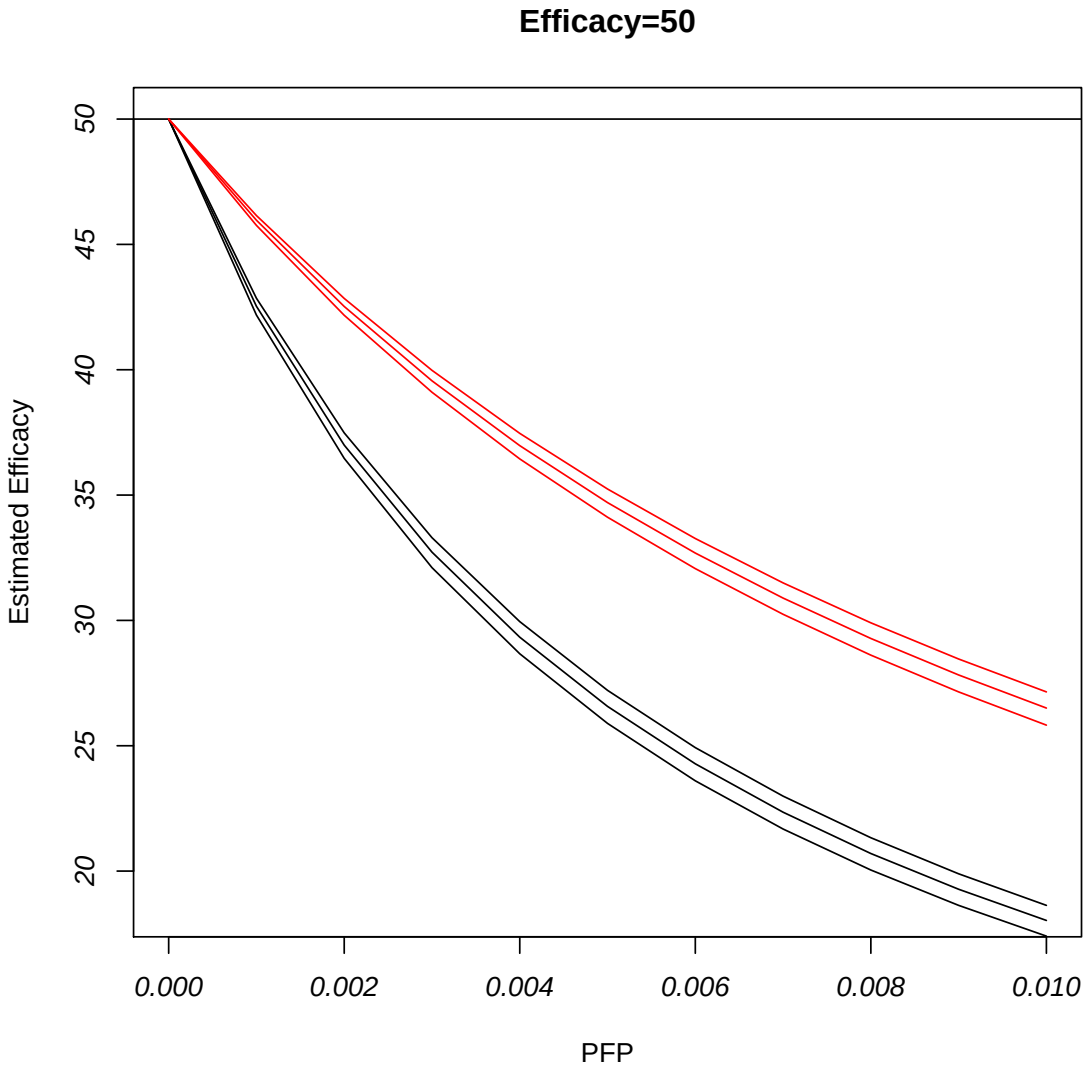

Figure 3: Illustration of bias due to misclassification at efficacy = 50% (solid horizontal line). Plot of what is estimated using the observed rates as a function of PFP. Black lines are for different PFNS based on $\pi_p = .006$ with $\pi_v = .0003$. Red lines are based on $\pi_p = .012$ with $\pi_v = .0006$.

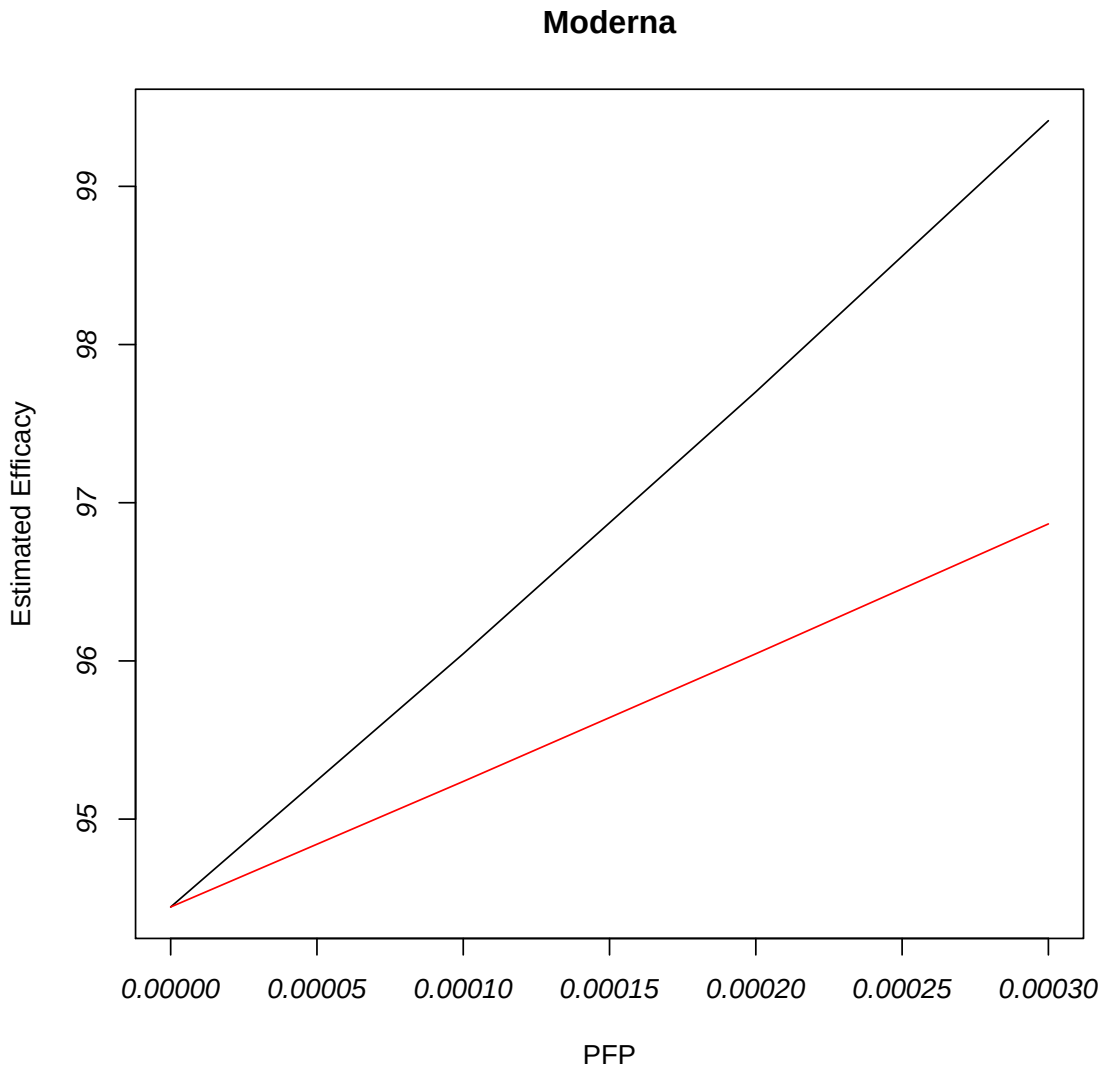

Figure 4: Illustration of correction of estimated efficacy with the Moderna data as a function of possible values of PFP. Black line uses the original data while red line is with doubled infection rates. Original estimated efficacy equals 94.4% in each case

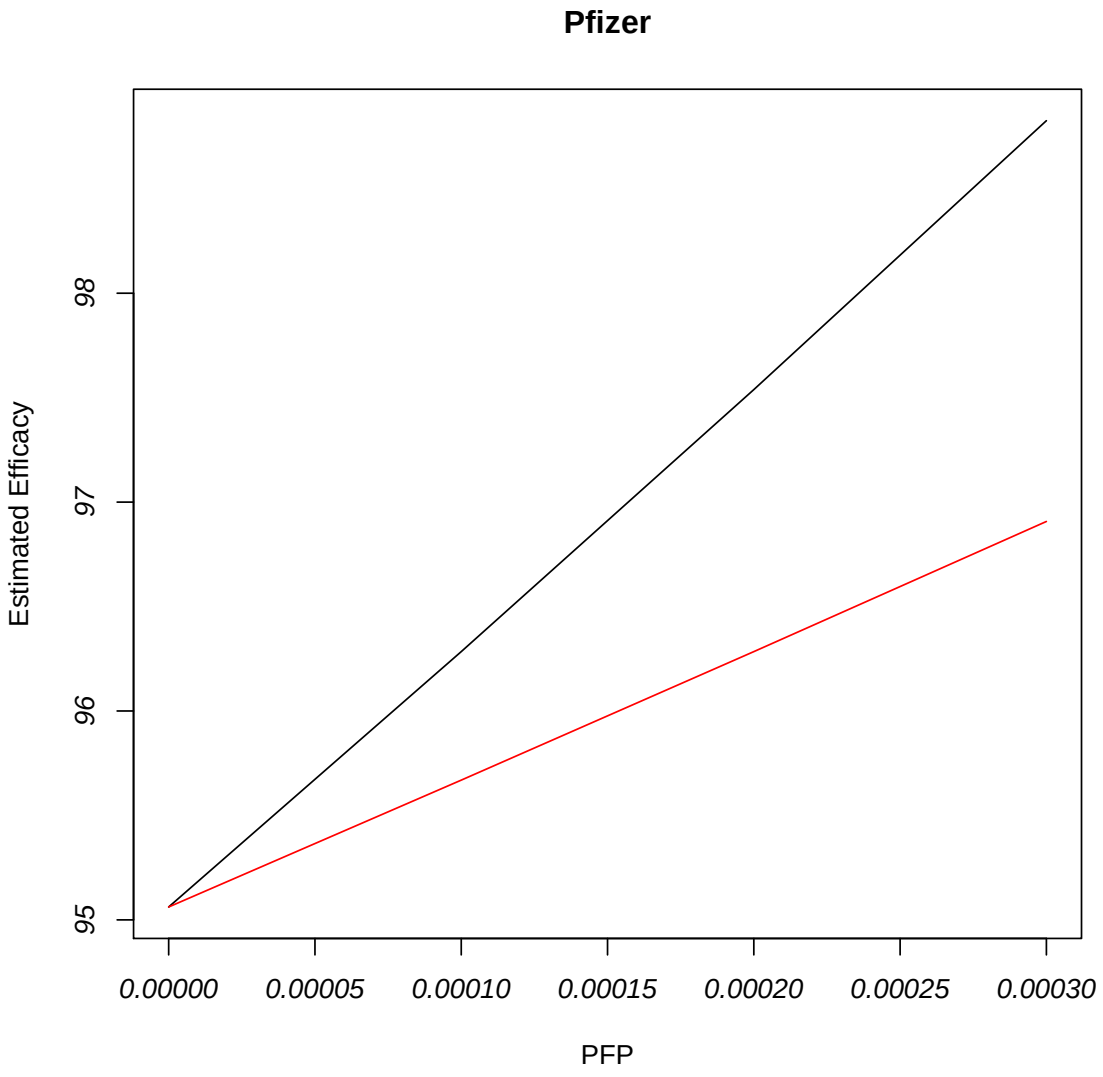

Figure 5: Illustration of correction of estimated efficacy with the Pfizer data as a function of possible values of PFP. Black line uses the original data while red line is with doubled infection rates. Original estimated efficacy equals 95.04% in each case

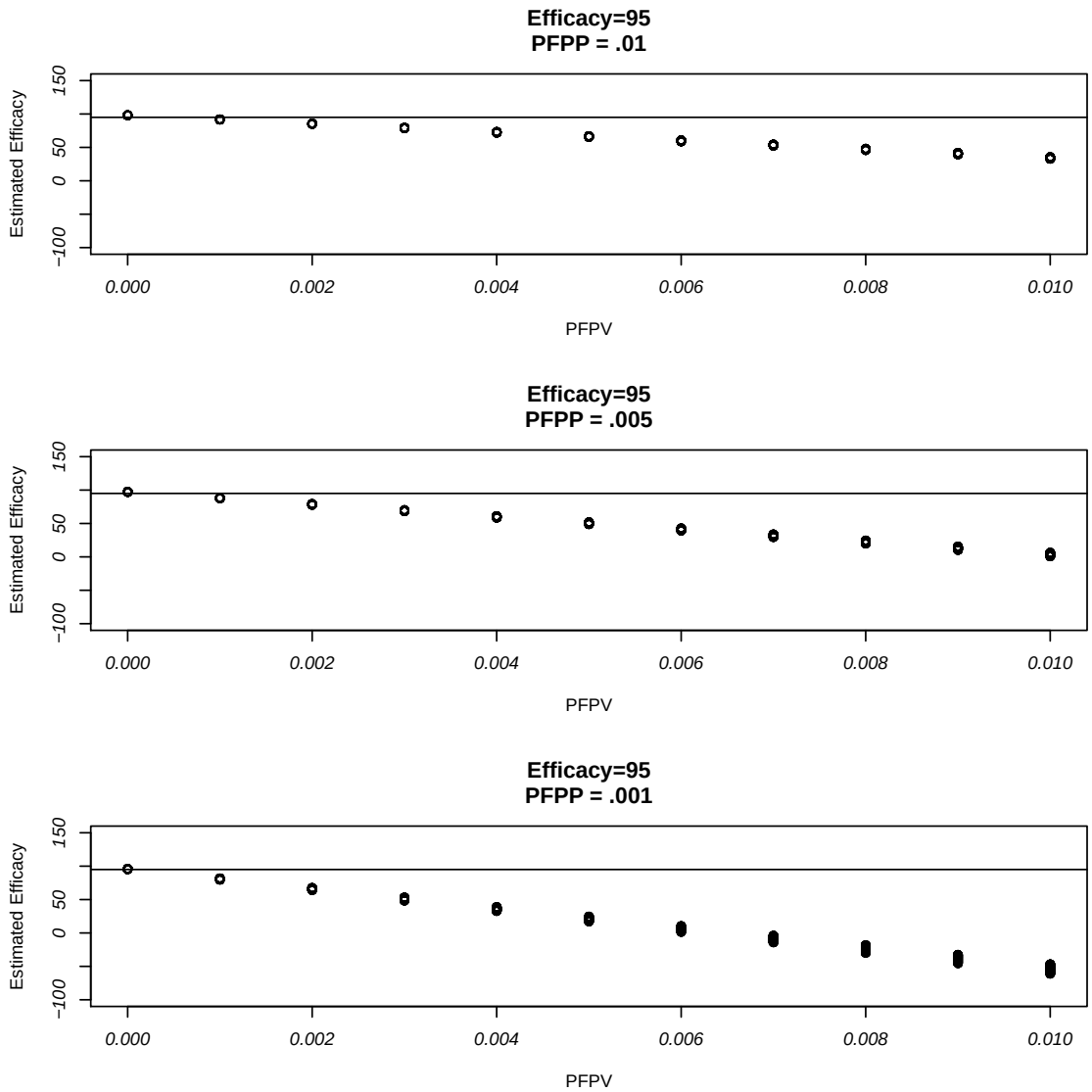

Figure 6: Illustration of bias due to differential misclassification at efficacy = 95% (solid horizontal line). PFPV = probability of false positive in vaccinated group. PFPP = probability of false positive in placebo group. Plot of what is estimated using the observed rates as a function of PFPV with different points from combinations of probabilities of false negatives over both groups. Based on $\pi_p = .006$ and $\pi_v = .0003$.

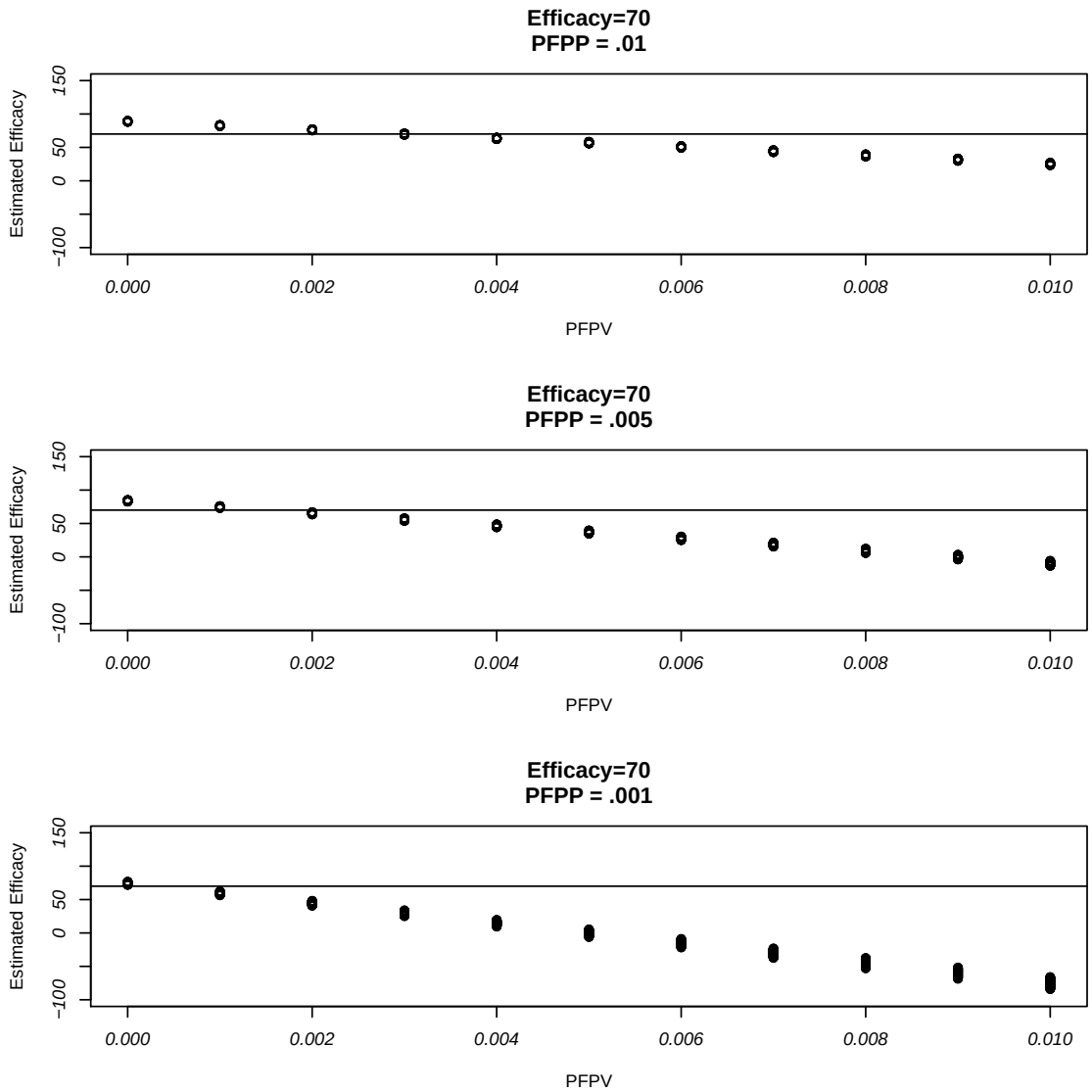

Figure 7: Illustration of bias due to differential misclassification at efficacy = 70% (solid horizontal line). PFPV = probability of false positive in vaccinated group. PFPP = probability of false positive in placebo group. Plot of what is estimated using the observed rates as a function of PFPV with different points from combinations of probabilities of false negatives over both groups. Based on $\pi_p = .006$ and $\pi_v = .0003$.

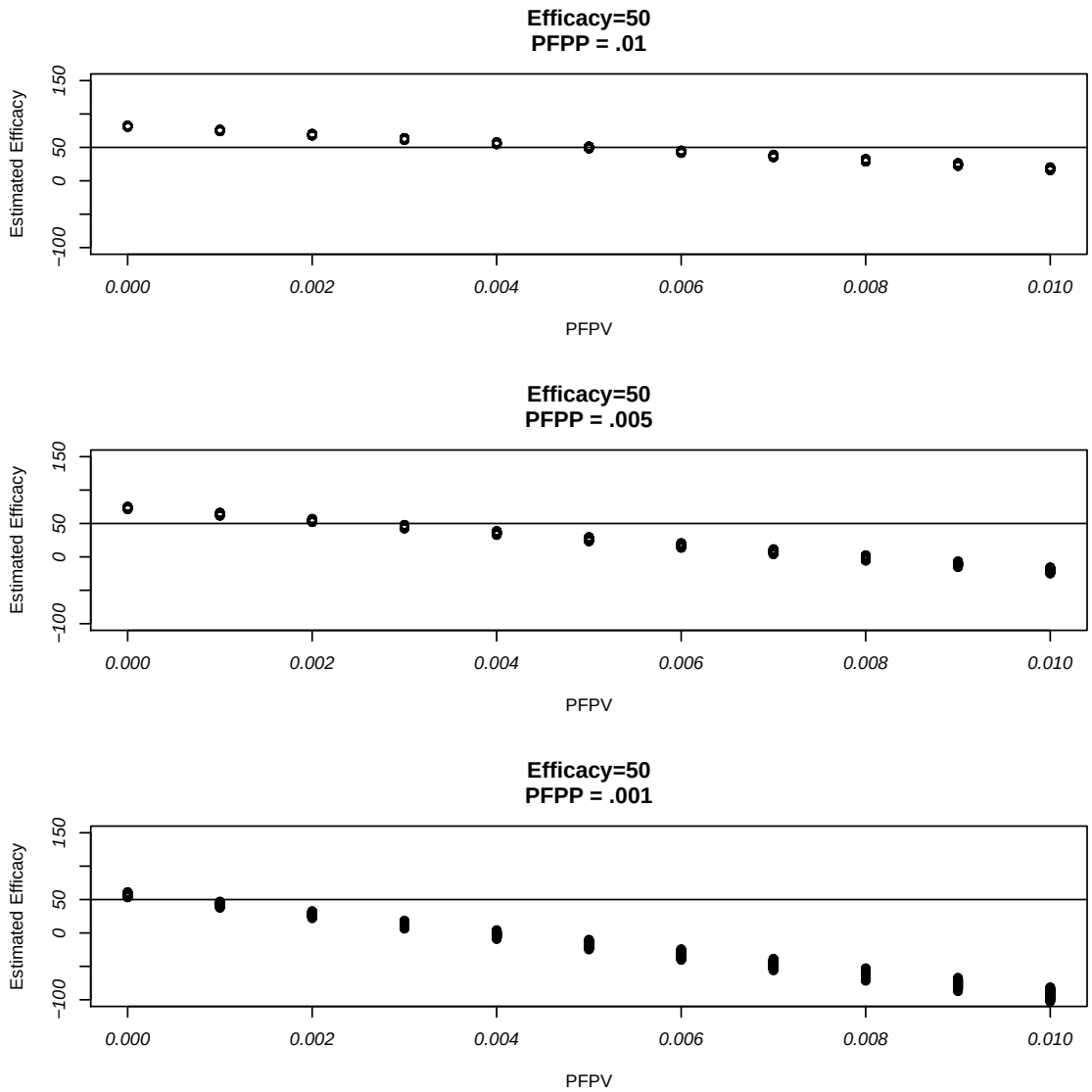

Figure 8: Illustration of bias due to differential misclassification at efficacy = 50% (solid horizontal line). PFPV = probability of false positive in vaccinated group. PFPP = probability of false positive in placebo group. Plot of what is estimated using the observed rates as a function of PFPV with different points from combinations of probabilities of false negatives over both groups. Based on $\pi_p = .006$ with $\pi_v = .0003$.

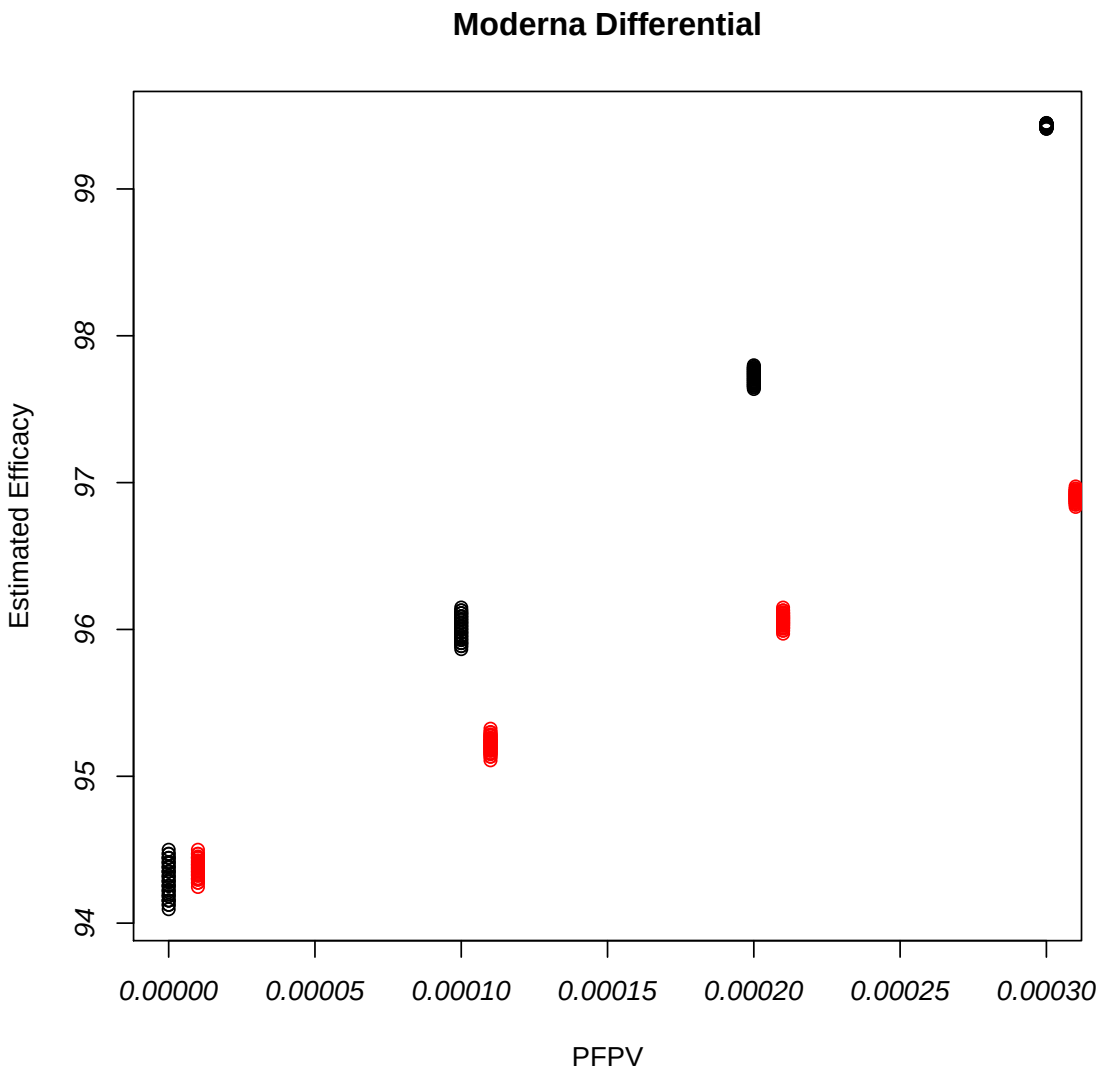

Figure 9: Illustration of corrected estimated efficacy with the Moderna data as a function of possible values of PFPV = probability of false positive in vaccinated group with differential misclassification. Black points are for the original data and range over all values of PFP in placebo groups and PFNs in the two groups. Red points (offset slightly for plotting purposes) use double the original infection rates. Original estimated efficacy equals 94.4% in each case.

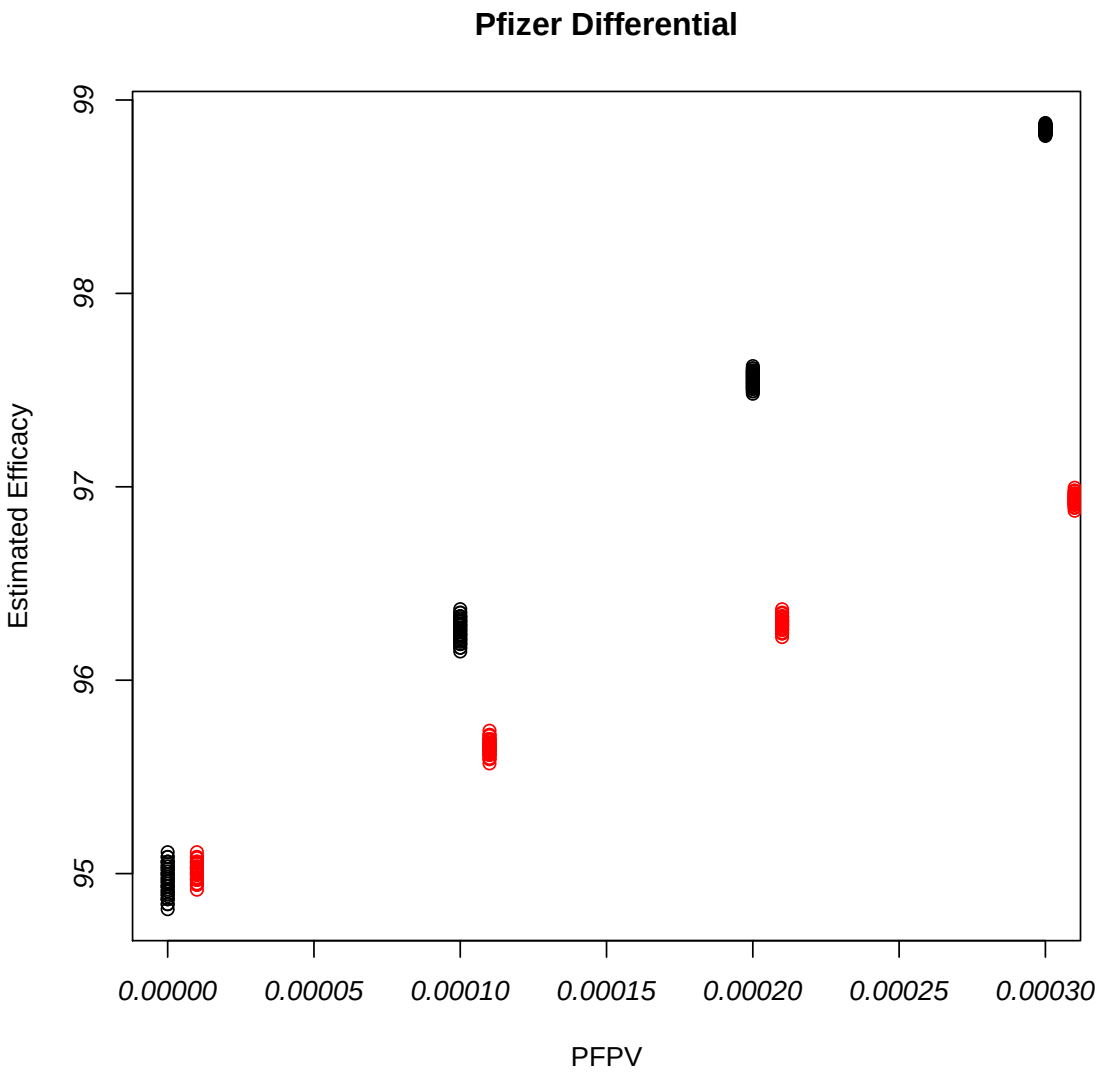

Figure 10: Illustration of corrected estimated efficacy with the Prizer data as a function of possible values of PFPV = probability of false positive in the vaccinated group with differential misclassification. Black points are for the original data and range over all values of PFP in placebo groups and PFNs in the two groups. Red points (offset slightly for plotting purposes) use double the original infection rates. Original estimated efficacy equals 95.04% in each case.