

Off-season VE estimates

November 12, 2019

1 Model

Assumptions:

1. All observations are independent.
2. Proportion p of the population is exposed to influenza (is at risk of disease of interest).
3. The unvaccinated have the probability f of infection with influenza.
4. The only differences between the off-season and the during-season populations are p and f .
5. Vaccination decreases the probability of infection by influenza by a factor $1 - e$ where $0 \leq e \leq 1$.
6. The entire population is at the same risk l of infection with other respiratory pathogens.
7. Vaccination does not modify the probability of infection with other pathogens.
8. No misclassification (investigated in Section 4).
9. The entire infected population (infected by influenza or another pathogen) is included into the test-negative study.

The expected proportions of cases and controls of a test-negative study under the above assumptions are in Table 1.

Table 1: Expected proportions of a test-negative study under the assumed model. F — infected with flu (cases). \bar{F} — infected with a pathogen other than flu (non-cases). V — vaccinated. \bar{V} — unvaccinated. f — probability of flu in unvaccinated. e — vaccine efficacy. v — vaccinated proportion. l — probability of infection by a non-influenza pathogen.

	F	\bar{F}
V	$pvf(1 - e)$	vl
\bar{V}	$p(1 - v)f$	$(1 - v)l$

The expected odds ratio of a test-negative study is

$$\text{OR}_{\text{TN}} = \frac{pvf(1 - e)(1 - v)l}{vlp(1 - v)f} = 1 - e \quad (1)$$

2 Low flu activity

The expected odds ratio of a test-negative study in Eq. 1 is unbiased and does not depend on the probability of flu infection f . This means that the test-negative VE estimate should be unbiased at any probability of flu infection. Figure 1 shows the results of 100,000 simulations of test-negative studies.

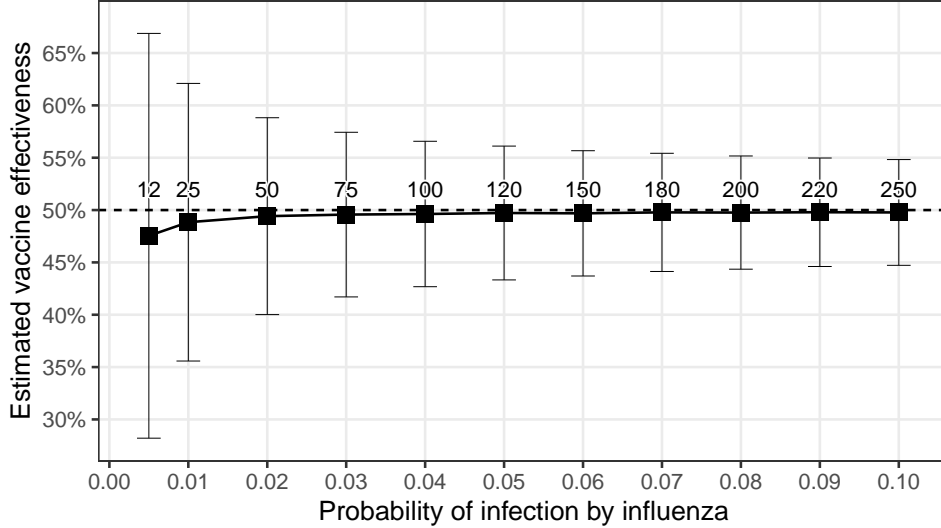


Figure 1: Estimated vaccine efficacy from 100,000 simulations of test-negative studies at different probabilities of flu infection. The true vaccine efficacy was 0.5 (indicated by the dashed line). The points represent probabilities at which the simulations were done. The error bars represent the expected standard deviation of the VE estimate. The numbers represent the expected number of cases in one study. The expected number of controls in all studies was 1,000. The deviation of the mean of the VE estimate across the simulations is negligible due to the large number of simulations. The starting population size was 10,000. Everyone in the population was at risk (proportion at risk was 1). The probability of non-influenza infection was 0.1. The vaccinated proportion was 0.5.

The results in Figure 1 show that the estimate remains almost unbiased even at very flu infection probabilities which result in low counts of cases in the test-negative studies. The variation of the estimate across studies predictably decreases with increasing flu infection probability due to the increasing number of cases.

There is a small ($<3\%$) amount of bias evident at $f \leq 0.01$. To see where this is coming from, consider Table 2 which defines notation for case/noncase counts of a test-negative study.

2 LOW FLU ACTIVITY

Table 2: Notation for counts of a test-negative study. F — infected with flu (cases). \bar{F} — infected with a pathogen other than flu (non-cases). V — vaccinated. \bar{V} — unvaccinated.

	F	\bar{F}
V	$X_{c,v}$	$X_{n,v}$
\bar{V}	$X_{c,u}$	$X_{n,u}$

Under the assumptions of the model, $E(X_{c,v}) = vf(1-e)$, $E(X_{c,u}) = (1-v)f$, $E(X_{n,v}) = vl$, $E(X_{n,u}) = (1-v)l$, and so $\frac{E(X_{c,v})E(X_{c,u})}{E(X_{n,v})E(X_{n,u})}$ is unbiased. However, this unbiased quantity is equivalent to the expected VE estimate when the individual counts are averages from many identical studies. The expected VE estimate from one study, $E(\frac{X_{c,v}X_{c,u}}{X_{n,v}X_{n,u}})$ is not the same quantity. The bias that results is related to the fact that the expected value of the reciprocal of a binomial random variable is not the reciprocal of its expected value.

$$E\left(\frac{1}{X}\right) \neq \frac{1}{E(X)} \quad X \sim \text{Binomial}(n, p) \quad (2)$$

This bias depends on the expected value, np , as shown in Figure 2 more so than it does on the choice of n and p (Figure 3).

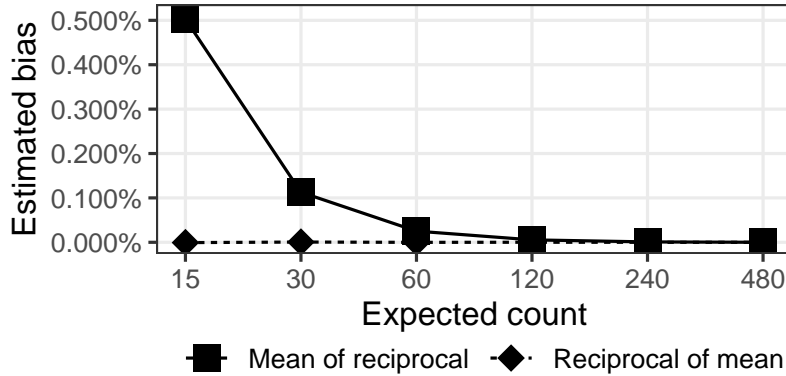


Figure 2: Results of simulating random numbers from $\text{Binomial}(n, p)$ where np was set to a value shown on the horizontal axis and n to 500. Points represent values for which 1,000,000 numbers were simulated. The vertical axis shows percent deviation from the true value. Reciprocal of mean means that 1 was divided by the mean taken across all generated numbers (the dashed line). Mean of reciprocal means that 1 was divided by each generated number and the mean taken across the resulting ratios (the solid line).

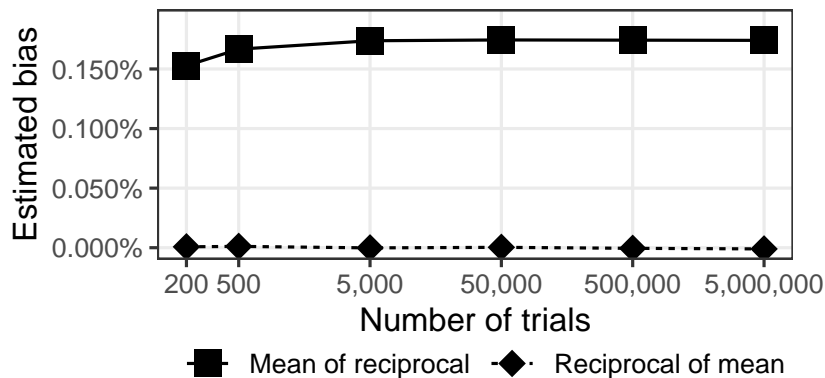


Figure 3: Results of simulating random numbers from $\text{Binomial}(n, p)$ where n was set to a value shown on the horizontal axis and np to 25. Points represent values for which 1,000,000 numbers were simulated. The vertical axis shows percent deviation from the true value. Reciprocal of mean means that 1 was divided by the mean taken across all generated numbers (the dashed line). Mean of reciprocal means that 1 was divided by each generated number and the mean taken across the resulting ratios (the solid line).

The only way to exaggerate this bias any more than it already has been in the simulations in Figure 1 would be to decrease the number of controls to very low values as well as cases. I ran an additional set of simulations where both f and l were set to 0.005 (while $p = 1$, $v = 0.5$, $e = 0.5$) with a starting population size of 10,000. The expected number of cases in these simulations was 37.5 and the expected number of controls was 50. The resulting bias was 4.5% toward the null. So even close to its maximum, this bias is very small and is unlikely to be important in real data.

3 Inclusion of unexposed population

If there is a subset of the population that is not exposed to the virus, the people in that subset are at no risk of being infected. Their inclusion into the study may raise concerns but as can be seen from Eq. 1, the odds ratio of a test-negative study does not depend on the proportion of the population “at risk”. Figure 4 summarises simulation results.

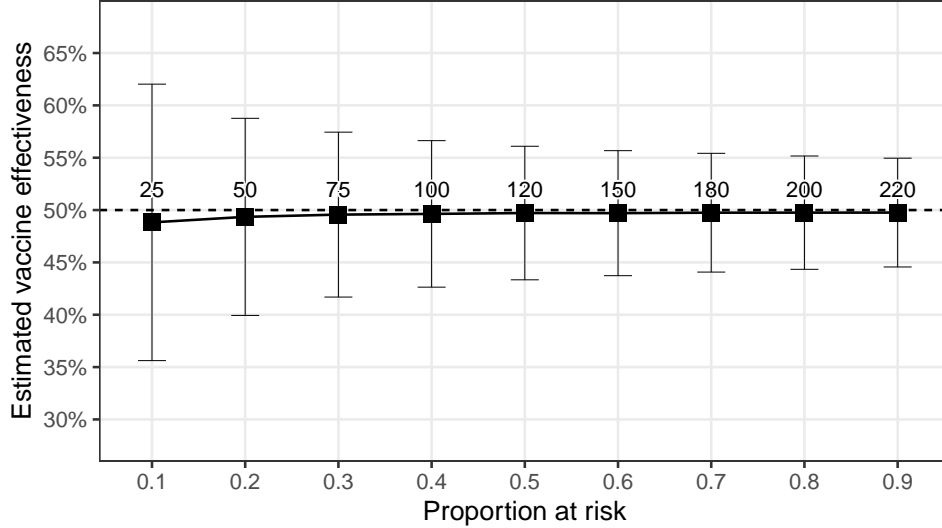


Figure 4: Estimated vaccine efficacy from 100,000 simulations of test-negative studies at different proportions of the sub-population at risk. The true vaccine efficacy was 0.5 (indicated by the dashed line). The points represent probabilities at which the simulations were done. The error bars represent the expected standard deviation of the VE estimate which predictably decreases with proportion at risk due to the increasing number of cases. The numbers represent the expected number of cases in one study. The expected number of controls in all studies was 1,000. The deviation of the mean of the VE estimate across the simulations is negligible due the large number of simulations. The starting population size was 10,000. The probability of flu was 0.1. The probability of non-influenza ARI was 0.1. The vaccinated proportion was 0.5.

The results in Figure 4 mimic those in Figure 1. Under the assumptions, decreasing the proportion at risk is equivalent to decreasing the probability of infection and is not expected to result in notable bias.

In general, the unexposed population will only contribute to the noncase counts. Table 3 defines notation for the expected counts of cases and noncases in the total, the exposed and the unexposed populations.

3 INCLUSION OF UNEXPOSED POPULATION

Table 3: Notation for counts of a test-negative study in total, exposed and unexposed populations. F — infected with flu (cases). \bar{F} — infected with a pathogen other than flu (non-cases). V — vaccinated. \bar{V} — unvaccinated.

	Exposed		Unexposed		Total	
	F	\bar{F}	F	\bar{F}	F	\bar{F}
V	n_1	n_3	0	n_5	n_1	$n_3 + n_5$
\bar{V}	n_2	n_4	0	n_6	n_2	$n_4 + n_6$

Under the assumptions, the unbiased odds ratio is $\frac{n_1 n_4}{n_2 n_3}$. The odds ratio in the general population is $\frac{n_1(n_4+n_6)}{n_2(n_3+n_5)}$.

As long as the vaccinated proportion is the same in the exposed and the unexposed populations, $\frac{n_3}{n_4} = \frac{n_5}{n_6} = v$ and

$$\frac{n_1(n_4 + n_6)}{n_2(n_3 + n_5)} = \frac{n_1(n_4 + n_6)}{n_2(vn_4 + vn_6)} = \frac{n_1}{n_2v} = \frac{n_1 n_4}{n_2 n_3} \quad (3)$$

Meaning that the expected odds ratio is the same in the total and the exposed populations.

4 Misclassification

Misclassification will introduce bias to both the during-season VE estimates and the off-season VE estimates. The off-season estimates may be affected more due to the smaller sample size.

To investigate, I ran the same simulations as reported in Figure 1 but in presence of non-differential misclassification described in Table 4. The results are in Figure 5.

Table 4: Parameters of non-differential missclassification used in the simulations.

Parameter	Value
Sensitivity of vaccination measurement	0.95
Specificity of vaccination measurement	0.8
Sensitivity of flu test	0.86
Specificity of flu test	0.984

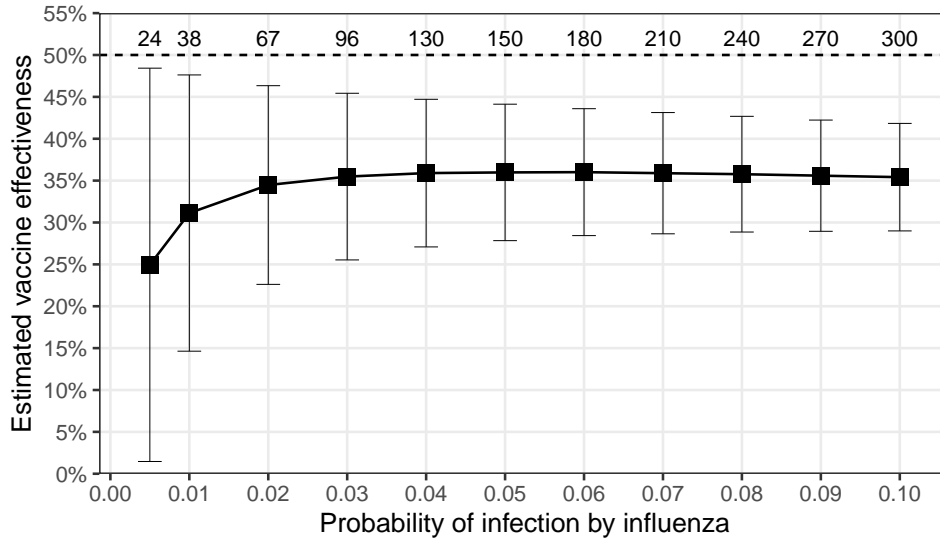


Figure 5: Estimated vaccine efficacy from 100,000 simulations of test-negative studies at different probabilities of flu infection. The true vaccine efficacy was 0.5 (indicated by the dashed line). The points represent probabilities at which the simulations were done. The error bars represent the expected standard deviation of the VE estimate. The numbers represent the expected number of cases in one study. The expected number of controls in all studies was 1,000. The deviation of the mean of the VE estimate across the simulations is negligible due the large number of simulations. The starting population size was 10,000. Everyone in the population was at risk (proportion at risk was 1). The probability of non-influenza infection was 0.1. The vaccinated proportion was 0.5. Parameters associated with misclassification are in Table 4.

5 CONCLUSION

The results show that the VE estimates are more biased in presence of missclassification. This applies to both the during-season (high flu activity) and the off-season (low flu activity) estimates. When flu activity is very low (<0.02), the VE estimates are affected to a greater extent due to the decreased number of cases.

I ran additional simulations of a very small study (probability of flu and non-flu both set to 0.005). The resulting mean VE estimate was 29% with a standard deviation of 35% which is fairly similar to the left-most point in Figure 5.

With a smaller number of cases, misclassification has a greater impact on the VE estimates. The smaller number of cases also makes the VE estimates more variable. So while greater bias may be expected from smaller studies (e.g. off-season), the off-season VE estimates cannot be expected to be consistently null (or close to null) under the model assumptions (one of which is that the vaccines still work off-season).

5 Conclusion

Neither the low flu activity nor the inclusion of the unexposed population can be expected to result in notable bias of VE estimates. If the VE estimates are consistently low when taken off-season, it cannot be explained by the fact that flu activity is low nor by the fact that much of the population is unexposed.