Retrospective, Observational Studies for Estimating Vaccine Effects on the Secondary Attack Rate of SARS-CoV-2

Marlena Bannick, Fei Gao, Elizabeth Brown, Holly Janes

Contents

Web Appendix 1

To illustrate the potential impact of event-prompted or infrequent testing on estimation of VE against transmission, we compare the target estimand with the actual estimands under imperfect testing. We assume a range of scenarios around SARS-CoV-2 infection and transmission that are intended to illustrate the issues of symptom-prompted testing and infrequent testing in isolation; in practice there may be several issues at play. We describe these scenarios below.

In the reference scenario, we assume that we have a prospectively defined group so that we correctly identify the primary case if they are tested, and any active infection is always identified if they are tested. For simplicity, all contacts are un-vaccinated and if infected the symptom presentation and duration of infection is unrelated to attributes of the primary case or transmission unit. Therefore, symptom-prompted, or infrequent testing of contacts does not change the estimand. Reflecting the high efficacy of COVID-19 vaccines against symptomatic infection, we fix vaccine efficacy against symptomatic infection at 0.9. Vaccine efficacy against SARS-CoV-2 infection is set at 0.5, so that the proportion of symptomatic infections in the vaccinated is 20% of that in the un-vaccinated. Given infection, the probability of symptoms among the un-vaccinated is 0.5, and the reduction in SAR for asymptomatic vs. symptomatic infections, and the overall VE against the SAR (the target estimand) are varied. Under infrequent testing, we assume that the duration of infection is uniformly distributed over 14 days, with a mean of 14 days for the un-vaccinated, and 8 days for the vaccinated. We vary the frequency of testing from once every 1 day to once every 14 days.

Web Appendix 2

Symptom-Prompted Testing

Let T be an indicator for transmission, S be an indicator for being symptomatic, and V be an indicator for vaccinated. Let μ represent the ratio of transmission probability between those vaccinated and un-vaccinated with $0 \leq \mu \leq 1$, i.e.,

$$P(T = 1|V = 1) = \mu P(T = 1|V = 0).$$

Vaccine efficacy is then defined as $1 - \mu$. We can rewrite each of the probabilities in terms of symptom development:

$$P(T = 1|V = 0) = P(T = 1|S = 1, V = 0)P(S = 1|V = 0) + P(T = 1|S = 0, V = 0)P(S = 0|V = 0)$$
$$P(T = 1|V = 1) = P(T = 1|S = 1, V = 1)P(S = 1|V = 1) + P(T = 1|S = 0, V = 1)P(S = 0|V = 1).$$

We now define the following five parameters, and we will rewrite the vaccine efficacy parameter $1 - \mu$ as a function of these parameters.

$$\lambda := \frac{P(S = 1|V = 1)}{P(S = 1|V = 0)}$$
$$\delta := \frac{P(T = 1|S = 0, V = v)}{P(T = 1|S = 1, V = v)}$$
$$\nu := \frac{P(T = 1|S = s, V = 1)}{P(T = 1|S = s, V = 0)}$$
$$\rho := P(S = 1|V = 0)$$
$$\tau := P(T = 1|S = 1, V = 0).$$

That is, λ represents the reduction in symptoms given vaccination, δ represents the reduction in transmission potential for asymptomatic individuals, holding vaccine status constant, and ν represents the reduction in transmission potential for vaccinated individuals, holding symptom presence constant. The probability of symptoms in an un-vaccinated person is ρ, and the probability of transmission for a symptomatic, un-vaccinated person is τ. Our goal is to solve for μ given fixed values of these quantities. Replacing the earlier

probabilities with these relationships,

$$\begin{aligned}
P(T=1|V=0) &= P(T=1|S=1,V=0)P(S=1|V=0) + P(T=1|S=0,V=0)P(S=0|V=0) \\
&= \rho P(T=1|S=1,V=0) + (1-\rho)P(T=1|S=0,V=0) \\
&= \rho P(T=1|S=1,V=0) + (1-\rho)\delta P(T=1|S=1,V=0) \\
&= \tau\big\{\rho + (1-\rho)\delta\big\} \\
P(T=1|V=1) &= P(T=1|S=1,V=1)P(S=1|V=1) + P(T=1|S=0,V=1)P(S=0|V=1) \\
&= P(T=1|S=1,V=1)P(S=1|V=1) + P(T=1|S=0,V=1)(1-P(S=1|V=1)) \\
&= P(T=1|S=1,V=1)\lambda\rho + P(T=1|S=0,V=1)(1-\lambda\rho) \\
&= P(T=1|S=1,V=1)\lambda\rho + P(T=1|S=1,V=1)\delta(1-\lambda\rho) \\
&= P(T=1|S=1,V=0)\nu\lambda\rho + P(T=1|S=1,V=0)\delta(1-\lambda\rho)\nu \\
&= \tau\nu\big\{\lambda\rho + \delta(1-\lambda\rho)\big\}.
\end{aligned}$$

Thus, using the relationship $P(T=1|V=1) = \mu P(T=1|V=0)$, we can find the vaccine efficacy μ:

$$\begin{aligned}
P(T=1|V=1) - \mu P(T=1|V=0) = \tau\nu\big\{\lambda\rho + \delta(1-\lambda\rho)\big\} - \mu\tau\big\{\rho + (1-\rho)\delta\big\} &\equiv 0 \\
\implies \nu\big\{\lambda\rho + \delta(1-\lambda\rho)\big\} &= \mu\big\{\rho + (1-\rho)\delta\big\} \\
\implies \mu &= \frac{\nu\big\{\lambda\rho + \delta(1-\lambda\rho)\big\}}{\rho + \delta(1-\rho)} \qquad (1)
\end{aligned}$$

With only symptomatic sampling of the index cases, we would instead have

$$P(T=1|S=1,V=1) - \tilde{\mu}P(T=1|S=1,V=0) = \tau\nu - \tau\tilde{\mu}$$

so, in this case, $\tilde{\mu} = \nu$, exactly the reduction in transmission given vaccination, holding symptom status constant. Therefore, the parameter being estimated when we only sample symptomatic index cases is the reduction in transmission with vaccination, only among cases that become symptomatic. If vaccination has no effect on development of symptoms, $\lambda = 1$, and we can see that (1) reduces to $\mu = \nu$, so the estimands μ and $\tilde{\mu}$ agree. The estimands also agree when $\nu = 0$, i.e., vaccinated individuals cannot transmit infection. When $\nu = 0$, then $\mu = \tilde{\mu} = 0$, i.e., the vaccine efficacy $1 - \mu$ is 1.

Web Appendix 3

Infrequent Testing

Let T be an indicator for transmission, N be a positive-valued random variable indicating duration of infection, and V be an indicator for vaccinated. As before, let μ represent the ratio of transmission probability between those vaccinated and un-vaccinated with $0 \leq \mu \leq 1$, i.e.,

$$P(T = 1|V = 1) = \mu P(T = 1|V = 0).$$

We can rewrite each of the probabilities in terms of duration of infectiousness:

$$P(T = 1|V = 0) = \int P(T = 1|N = n, V = 0) dP_{N|V=0}(n)$$
$$P(T = 1|V = 1) = \int P(T = 1|N = n, V = 1) dP_{N|V=1}(n).$$

Let $P_{N|V=0}(n)$ and $P_{N|V=1}(n)$ be the distribution function of infection duration for infected people who are un-vaccinated and vaccinated, respectively. Let τ_0 be the daily risk of transmission to a contact among the un-vaccinated and $\tau_1 = \tau_0\nu$, $\nu \leq 1$ be the daily risk of transmission to a contact among the vaccinated. Define $\nu = \tau_1/\tau_0 \leq 1$. We assume that there is a constant risk of infection across the duration of infection ($P(T = 1|N = n, V = v) = 1 - \exp(-n\tau_v) \approx n\tau_v$ for small τ_v). Then,

$$P(T = 1|V = 0) = \int n\tau_0 dP_{N|V=0}(n)$$
$$P(T = 1|V = 1) = \int n\tau_1 dP_{N|V=1}(n).$$

If we assume that $P_{N|V=0}(n) = \text{Unif}(\rho_0 - c, \rho_0 + c)$, $P_{N|V=1}(n) = \text{Unif}(\rho_1 - c, \rho_1 + c)$, and with $\lambda = \rho_1/\rho_0 \leq 1$ (and in particular, to have a sensible minimum duration of infection among the vaccinated [i.e., $\rho_1 - c$], we need to have $\lambda > c/\rho_0$), then we have,

$$\begin{aligned} P(T = 1|V = 0) &= \tau_0 \, \mathbb{E}_{N|V=0}\{N\} \\ &= \tau_0\rho_0 \\ P(T = 1|V = 1) &= \tau_1 \, \mathbb{E}_{N|V=1}\{N\} \\ &= \tau_1\rho_1 \\ &= \tau_0\rho_0\lambda\nu. \end{aligned}$$

Therefore, we have $\mu = \lambda\nu$.

Let S be an indicator of the index case being included in the sample. Let k be the window of time during which an infection may be sampled (e.g., $k = 14$ means there is a two-week period when someone is infected during which a single test can be taken). In this testing scenario we consider $P(S = 1|N = n) = \max(n/k, 1)$, i.e., the probability of being sampled is proportional to the duration of infection, and it is dependent on no other variables. For fixed vaccine status (where we note that with $P_{N|V=v}$ defined above, $\mathrm{Var}_{N|V=v}\{N\} = c^2/3$), we have:

$$\begin{aligned}
P(T = 1|V = v, S = 1) &= \int P(T = 1|V = v, S = 1, N = n)P(N = n, |V = v, S = 1)dn \\
&= \int P(T = 1|V = v, N = n)P(N = n|V = v, S = 1)dn \quad \text{since } T \perp S|N \\
&= \int n\tau_v \frac{P(S = 1|N = n, V = v)P(N = n|V = v)}{P(S = 1, V = v)}dn \quad \text{by Bayes formula} \\
&= \tau_v \int n \frac{P(S = 1|N = n)}{\int P(S = 1|N = n, V = v)dP_{N|V=v}(n)}dP_{N|V=v}(n) \quad \text{since } S \perp V|T \\
&= \tau_v \int n \frac{P(S = 1|N = n)}{\int P(S = 1|N = n)dP_{N|V=v}(n)}dP_{N|V=v}(n) \quad \text{since } S \perp V|T
\end{aligned}$$

We first consider the case where $k > \rho + c$, i.e., no one has a probability of testing greater than 1. In that case, we can have only one integral:

$$\begin{aligned}
&= \tau_v \int n\{n/k\} \frac{1}{P(S = 1|V = v)}dP_{N|V=v} \\
&= \tau_v \int n\{n/k\} \frac{1}{\int P(S = 1|N = n, V = v)dP_{N|V=v}}dP_{N|V=v} \\
&= \tau_v \int n^2(1/k) \frac{1}{\int (n/k)dP_{N|V=v}}dP_{N|V=v} \\
&= \tau_v \, \mathbb{E}_{N|V=v}\{N^2\}/\, \mathbb{E}_{N|V=v}\{N\} \\
&= \frac{\tau_v}{\rho_v}\{\, \mathbb{E}_{N|V=v}\{N\}^2 + \mathrm{Var}_{N|V=v}\{N\}\} \\
&= \frac{\tau_v}{\rho_v}\{\rho_v^2 + (1/3)c^2\}. \qquad (2)
\end{aligned}$$

Interestingly, the parameter for the testing window size k drops out of the expression. Thus, the inferred

parameter for μ, $\tilde{\mu}$, when $k > \rho_0$ (which implies $k > \rho_1$ as well) is given by

$$\begin{aligned}\tilde{\mu} &= \frac{P(T=1|V=1,S=1)}{P(T=1|V=0,S=1)} \\ &= \frac{\tau_1/\tau_0}{\rho_1/\rho_0}\left\{\frac{\rho_1^2+(1/3)c^2}{\rho_0^2+(1/3)c^2}\right\} \\ &= \frac{\nu}{\lambda}\left\{\frac{\lambda^2\rho_0^2+(1/3)c^2}{\rho_0^2+(1/3)c^2}\right\}.\end{aligned}$$

Now consider the setting where $k \in [\rho_v - c, \rho_v + c]$. In this case people with duration $n \geq k$ will be sampled with probability 1. The denominator in the probability is $\int P(S=1|N=n)dP_{N|V=v}(n)$, which can be evaluated as (we use dP for simplicity):

$$\begin{aligned}\int P(S=1|N=n)dP_{N|V=v}(n) &= \int P(S=1|N=n)1(n\leq k)dP(n) + \int P(S=1|N=n)1(n>k)dP(n) \\ &= \int_{\rho_v-c}^{k}(n/k)dP(n) + \int_{k}^{\rho_v+c}dP(n) \\ &= \frac{1}{2ck}\Big\{\int_{\rho_v-c}^{k}(n)dn\Big\} + \frac{1}{2c}\Big\{\int_{k}^{\rho_v+c}dn\Big\} \\ &= \frac{1}{2ck}\left\{\frac{k^2-(\rho_v-c)^2}{2}\right\} + \frac{1}{2c}\Big\{(\rho_v+c)-k\Big\} \\ &= \frac{1}{4ck}\Big\{k^2-(\rho_v-c)^2+2k(\rho_v+c)-2k^2\Big\} \\ &= \frac{1}{4ck}\Big\{2k(\rho_v+c)-(\rho_v-c)^2-k^2\Big\} =: S_k.\end{aligned}$$

Making sure that this still equals ρ_v/k, we have when $k = (\rho_v + c)$ that:

$$\begin{aligned}\int P(S=1|N=n)dP_{N|V=v}(n) &= \frac{1}{4c(\rho_v+c)}\Big\{2(\rho_v+c)^2-(\rho_v-c)^2-(\rho_v+c)^2\Big\} \\ &= \frac{1}{4c(\rho_v+c)}\Big\{(\rho_v+c)^2-(\rho_v-c)^2\Big\} \\ &= \frac{\rho_v^2+2\rho_v c+c^2-\rho_v^2+2\rho_v c-c^2}{4c(\rho_v+c)} \\ &= \frac{4\rho_v c}{4c(\rho_v+c)} \\ &= \frac{\rho_v}{\rho_v+c}.\end{aligned}$$

So, this is equivalent to what we got for the denominator expression in the first derivation when $n < k$. We now derive the full expression for $P(T=1|V=v,S=1)$ and check to make sure that we get the same thing as in (2) when $k = (\rho_v + c)$.

Putting this together with the numerator we have:

$$\begin{aligned}
P(T=1|V=v,S=1) &= \frac{\tau_v}{S_k}\Big\{\int n(n/k)1(n\le k)dP(n) + \int n1(n>k)dP(n)\Big\} \\
&= \frac{\tau_v}{S_k}\Big\{\frac{1}{k}\int_{\rho_v-c}^{k} n^2 dP(n) + \int_{k}^{\rho_v+c} ndP(n)\Big\} \\
&= \frac{\tau_v}{S_k}\Big\{\frac{1}{k}\int_{\rho_v-c}^{k} n^2 dP(n) + \int_{k}^{\rho_v+c} ndP(n)\Big\} \\
&= \frac{\tau_v}{S_k 2c}\Big\{\frac{1}{k}\int_{\rho_v-c}^{k} n^2 dn + \int_{k}^{\rho_v+c} ndn\Big\} \\
&= \frac{\tau_v}{S_k 2c}\left\{\frac{1}{3k}\{k^3-(\rho_v-c)^3\} + \frac{(\rho_v+c)^2-k^2}{2}\right\} \\
&= \frac{\tau_v}{12ckS_k}\Big\{2k^3 - 2(\rho_v-c)^3 + 3k(\rho_v+c)^2 - 3k^3\Big\} \\
&= \frac{\tau_v}{12ckS_k}\Big\{3k(\rho_v+c)^2 - k^3 - 2(\rho_v-c)^3\Big\}.
\end{aligned}$$

When $k=(\rho_v+c)$,

$$\begin{aligned}
P(T=1|V=v,S=1) &= \frac{\tau_v(\rho_v+c)}{12c\rho_v(\rho_v+c)}\Big\{3(\rho_v+c)^3 - (\rho_v+c)^3 - 2(\rho_v-c)^3\Big\} \\
&= \frac{\tau_v}{12c\rho_v}\Big\{2(\rho_v+c)^3 - 2(\rho_v-c)^3\Big\} \\
&= \frac{\tau_v}{6c\rho_v}\Big\{(\rho_v+c)^3 - (\rho_v-c)^3\Big\} \\
&= \frac{\tau_v}{6c\rho_v}\Big\{\rho_v^3 + 3\rho_v^2 c + 3\rho_v c^2 + c^3 - \rho_v^3 + 3\rho_v^2 c - 3\rho_v c^2 + c^3\Big\} \\
&= \frac{\tau_v}{6c\rho_v}\Big\{6\rho_v^2 c + 2c^3\Big\} \\
&= \frac{\tau_v}{\rho_v}\{\rho_v^2 + (1/3)c^2\}.
\end{aligned}$$

A third scenario is where $k<\rho_v-c$, such that all infections are captured with probability 1. In this case, $\int P(S=1|N=n)dP(n)=1$, and $P(T=1|V=v,S=1)=\tau_v\int ndP(n)=\tau_v\rho_v$. Therefore,

$$\begin{aligned}
\tilde{\mu} &= \frac{P(T=1|V=1,S=1)}{P(T=1|V=0,S=1)} \\
&= \frac{\tilde{\mu}_1}{\tilde{\mu}_0}
\end{aligned}$$

where

$$\tilde{\mu}_v = \tau_v\left\{\rho_v 1(k<\rho_v-c) + \big\{\rho_v + (1/3\rho_v)c^2\big\}1(\rho_v-c\le k\le\rho_v+c) + \frac{3k(\rho_v+c)^2 - k^3 - 2(\rho_v-c)^3}{12ckS_k}1(k>\rho_v+c)\right\}.$$

Importantly, $\tilde{\mu}_1$ and $\tilde{\mu}_0$ may differ with respect to which of the indicator functions is activated, since ρ_1 will be different than ρ_0 when $\lambda < 1$. Notice that if $k > \max(\rho_0, \rho_1)$, then the ratio of $\tilde{\mu}_1/\tilde{\mu}_0$ does not depend on k. There is still a difference between $\tilde{\mu}$ and μ, but it does not change with the interval k. We can see this relationship in Figure A1.

Web Figure 1: Relationship between the inferred vaccine efficacy parameter $1 - \tilde{\mu}$ and k, the testing interval. Colors indicate the true vaccine efficacy, which is also obtained by setting the testing interval to 1 day.