2022

On Efficacy and Effectiveness of Vaccines: A Mathematical Approach Based on Conditional Probability with Applications to the COVID-19 Context

Flavius Guias

*Dortmund University of Applied Sciences and Arts, flavius.guias@fh-dortmund.de*

Follow this and additional works at: https://ir.library.illinoisstate.edu/spora

Part of the Epidemiology Commons, Influenza Virus Vaccines Commons, and the Other Applied Mathematics Commons

**Recommended Citation**


DOI: https://doi.org/10.30707/SPORA8.1.1652717156.639924

Available at: https://ir.library.illinoisstate.edu/spora/vol8/iss1/7

This Mathematics Research is brought to you for free and open access by ISU ReD: Research and eData. It has been accepted for inclusion in Spora: A Journal of Biomathematics by an authorized editor of ISU ReD: Research and eData. For more information, please contact ISURed@ilstu.edu.
On Efficacy and Effectiveness of Vaccines: A Mathematical Approach Based on Conditional Probability with Applications to the COVID-19 Context

Flavius Guiaş

1 Introduction

The notions of efficacy and effectiveness related to vaccines have in the literature slightly different meanings, since researchers were always aware of the fact that the values computed by a the same formula in clinical trials and/or under ideal conditions differ in most cases from the values based on “real-life” data [3]. The normal approach in order to estimate the quality of a vaccine is to compute the relative risk in the vaccinated compared to the unvaccinated group. Therefore, the vaccine effectiveness (or observed effectiveness) $VE$ is defined by the formula

$$VE = 1 - \frac{P(I \mid V)}{P(I \mid V')},$$

with the obvious notations: $I$ for “infected,” $V$ for “vaccinated,” and $V'$ for “unvaccinated.” Applying Bayes’ formula we obtain

$$VE = 1 - \frac{P(V \mid I)P(I)}{P(V')} \cdot \frac{P(V')}{P(V)} \cdot \frac{P(V)}{P(V')P(I)} \cdot \frac{P(V')P(I)}{P(V | I)P(I)}$$

$$= 1 - \frac{P(V \mid I)}{P(V')} \cdot \frac{P(V')}{P(V)}$$

$$= 1 - \frac{P(V \mid I)}{1 - P(V \mid I)} \cdot \frac{P(V)}{1 - P(V)}$$

$$= \frac{P(V)}{1 - P(V \mid I)}.$$  

This formula is used for computing the vaccine effectiveness under “real life” conditions [2], the main reason being that the involved probabilities can be easily estimated from the observed statistical data. Nevertheless, these data may be subject to several bias sources. One possibility is the misclassification of the vaccination status [1]. The statistics of the COVID-19 pandemic show also a significant bias of another type. For example, the data made available by the weekly reports of the German RKI institute indicate a substantial decline in the effectiveness related to symptomatic infections [5, p. 25]. The explanation given is that the vaccine was developed for an ancient variant of the virus, while now the delta mutation prevails. But we will see that the main cause is the asymmetric exposure to the virus of the vaccinated and unvaccinated group. This fact explains also the apparent paradox illustrated by the UKHSA Vaccine Surveillance Reports from November 2021 [2, p. 33]. Except for the age group under 18, where the vaccination rate is the smallest, in all other cohorts the incidence (as positive tests, not only symptomatic cases) among vaccinated individuals is larger than that among unvaccinated of the same age group. This is explained in the report by the different behaviour of the groups of vaccinated and unvaccinated individuals, for example unvaccinated are more cautious than vaccinated, which may feel more safe and behave “normally,” or by the natural immunity acquired by a...
part of the former group. Another explanation for reduced infection rates among unvaccinated, perhaps the most important one, can be given by the restrictions imposed upon them, for example prohibition to attend public gatherings, gastronomy, entertainment, shopping, etc.

This means that being vaccinated doesn’t necessarily come along with a higher risk of being infected with SARS-CoV2, since the formulas (1), (3) can be biased by the possible dependence of vaccination and exposure to the virus. If the bias is very strong, their values are in fact meaningless for estimating the quality of a vaccine.

The main result of this paper gives a precise relation which is able to explain such facts. The organization is as follows.

In Section 2 we define the notions of efficacy and degree of discrimination and give a motivation for this choice. Briefly stated, in our framework, motivated mainly by the COVID-19 pandemic, where the probabilities of exposure of vaccinated and unvaccinated individuals can be highly different, a more relevant measure for the quality of a vaccine would be to take all probabilities in the formulas (1), (3) as conditional ones with respect to exposure to the virus, i.e. exposure to the virus, \( P_E(\cdot) \) or \( P(\cdot | E) \), and define the result as the efficacy denoted by \( VE^* \). We will show that this definition is consistent with the usual one used in epidemiology. In well designed clinical trials, the exposure probability should be independent on the vaccination status (otherwise, the crude statistical data would be not relevant). In this case we will show that under this assumption of independence, the efficacy estimated from such trials coincides with our definition using conditional probabilities.

In the context of clinical trials described above, it is obvious that no discrimination takes place among the groups of vaccinated and unvaccinated. However, in the real life, especially in the COVID-19 context, things turn out to be different. We define therefore a quantity called degree of discrimination. It is computed by a similar formula to (1), by replacing \( I \) with \( E \) and switching \( V \) and \( V' \). This number, given by “1 – the relative risk of exposure of the unvaccinated group compared to the vaccinated group,” quantifies the discrepancy in exposure to the virus of the vaccinated and unvaccinated population.

The main result of the paper is presented in Section 3 and consists in quantifying the relation between effectiveness, efficacy and degree of discrimination. We also discuss its possible consequences.

In Section 4 we perform similar considerations regarding the vaccine effectiveness regarding hospitalization, while in Section 5 we illustrate the theoretical results by examples related to the COVID-19 pandemic in Germany and the UK. We consider statistical data from November 2021, before the emergence of the omicron variant of SARS-CoV-2.

2 Efficacy and Degree of Discrimination

We next define the terms which are used in this paper.

Definition 1. The vaccine efficacy \( VE^* \) is defined as

\[
VE^* = 1 - \frac{P_E(I | V)}{P_E(I | V')}.
\]  

We point out that this definition is consistent with the standard usage of the term “efficacy” in epidemiology. Indeed, this notion is associated with clinical trials based on appropriate designs, where the assumption of independence of exposure and vaccination status can be made. If we assume that the trials are performed before a massive vaccination campaign of the population is started, then usually one compares two equally sized groups which are also statistically homogeneous, meaning that \( P(V) = P(V') = 0.5 \) and \( P(V \cap E) = P(V' \cap E) \), that is, the number of exposed individuals in each group should be in principle the same. In fact, this implies the independence of \( V \) and \( E \): Since \( P(E) = P(V \cap E) + P(V' \cap E) \), this means that \( P(V \cap E) = P(V' \cap E) = 0.5 \cdot P(E) = P(V) \cdot P(E) = P(V') \cdot P(E) \).

We finally note that the conditional probabilities involved in computing the efficacy from Definition 1 rely on information on exposure, which in practice is difficult to collect. An instrument for estimating them is a stratified analysis according to variables that may correlate with exposure to infection [4].

Definition 2. The degree of discrimination \( D \) is defined as

\[
D = 1 - \frac{P(E | V')}{P(E | V)}.
\]  

This term is a measure for the discrepancy in exposure to the virus between vaccinated and unvaccinated individuals. In the COVID-19 context, the additional restrictions imposed only on the latter, together with their more risk-averse behaviour, lead to a decrease the exposure probability \( P(E | V') \) compared to the value \( P(E | V) \) for the vaccinated population. Therefore in this paper we make the assumption \( D \geq 0 \). The maximal value is \( D = 1 \), which is attained if unvaccinated individuals are not exposed to the virus at all, meaning the complete isolation of them.

3 The Relation Between Effectiveness, Efficacy, and Degree of Discrimination

In this section we state the main result of this paper and discuss possible consequences.
Theorem 1. Between the quantities $VE$, $VE^*$, and $D$ the relation
\[
1 - VE^* = (1 - VE) \cdot (1 - D)
\] holds.

Proof. Conditioning by the event $E$, the exposure to the virus during a given time interval, we compute the relative risk as
\[
P_E(I \mid V) = \frac{P_E(V \cap I)}{P_E(V')} \cdot \frac{P_E(V')}{P_E(V')} = \frac{P(V \cap I \cap E)}{P(E)} \cdot \frac{P(E)}{P(I)} \cdot \frac{P(V')}{P(E)}
\]
and since $I \subseteq E$:
\[
P_E(I \mid V) = \frac{P(V \cap I)}{P(V' \cap I)} \cdot \frac{P_E(V')}{P_E(V')}
\]
Moreover, from (9) we obtain
\[
P(V \mid I) \cdot P(I) = P(V \cap I) = P(V' \cap I) \cdot P_E(V')
\]
and
\[
P(V' \mid I) = P(V' \cap I) \cdot P_E(V')
\]
Before going on with the computation, we point out the correspondence to the concepts from [4], where the vaccinated and unvaccinated status are denoted by 1 and 0, respectively. Probabilities conditioned on exposure to the virus are called transmission probabilities. The conditional probability $P_E(I \mid V)$ is the transmission probability $p_1$ to the vaccinated susceptibles conditioned by exposure due to contact to infected individuals of any vaccinated status. The dot means that the infective contact may have any status 0 or 1, otherwise transmission probabilities can be also differentiated according to the vaccination status. Similarly we have that $P_E(I \mid V')$ is the transmission probability $p_0$ to the unvaccinated susceptibles.

Moreover, from [9] we obtain
\[
P(V \cap I) \cdot P(V') = P(V \cap I) \cdot P(V' \cap I) \cdot P(V')
\]
that is, the relative risk computed here equals to the ratio of the secondary attack rates of the vaccinated and unvaccinated susceptibles respectively.

By Bayes’ formula we obtain the following from (10):
\[
P_E(I \mid V) = \frac{P(V \mid I)}{P(V')} \cdot \frac{P(E \mid V') \cdot P(V')}{P(E)} \cdot \frac{P(E)}{P(V)} = (1 - VE) \cdot \frac{P(E \mid V')}{P(E)}
\]
where in the last step formula (3) was used.

Taking into account Definitions 1 and 2 this is equivalent to $1 - VE^* = (1 - VE) \cdot (1 - D)$, which is exactly the statement of the theorem.

In the following we will discuss some consequences of this result.

In the present paper, under the assumption $D \geq 0$, from [8] can be easily seen that $VE^* \geq VE$, that is, the efficacy is always larger than the observed effectiveness.

Moreover, the result of Theorem 1 can be also stated as
\[
VE = 1 - \frac{1 - VE^*}{1 - D}.
\]

The above formula implies that for $D \geq VE^*$ we have $VE \leq 0$, i.e. a negative observed effectiveness if the degree of discrimination is larger than the efficacy. This may happen especially if we consider the efficacy regarding all types of infection, not only symptomatic. In this case its value is much lower than regarding only symptomatic infections and the degree of discrimination is likely to surpass it. This gives a mathematical explanation for the negative observed effectiveness in the UK [6, p.33] since its value is computed regarding all kind of infections, not only symptomatic. But this does not mean at all that the vaccines do not protect. The probability of infection of the vaccinated individuals is in fact higher due to their higher exposure, implied by a higher degree of discrimination, but the efficacy of the vaccines is always larger than the observed effectiveness.

4 Real Effectiveness Regarding Hospitalization

In contrast to the values of the observed effectiveness regarding the protection against all infections (not only symptomatic), which can be relatively low, the data related to the protection of the COVID-19 vaccines against severe cases, which may lead to hospitalization, treatment in intensive care or death still show high values. In this section we will discuss and explain this fact within the framework introduced previously, considering exemplarily the event of hospitalization, denoted with $H$.

Definition 3. a) We define the observed effectiveness regarding hospitalization $VE_H$ similarly to (1) or (2):
\[
VE_H = 1 - \frac{P(H \mid V)}{P(H \mid V')} = 1 - \frac{P(V \mid H)}{P(V' \mid H)} \cdot \frac{P(V')}{P(V)}.
\]

b) Similarly to (6) and (10) we define the real effectiveness regarding hospitalization:
\[
VE^*_H = 1 - \frac{P(H \mid V)}{P_H(H \mid V')} = 1 - \frac{P(V \mid H)}{P(V' \mid H)} \cdot \frac{P(V')}{P(V \mid I)}.
\]
For the real effectiveness regarding hospitalization it is therefore meaningful to take conditional probabilities with respect to $I$ that is, we use the relative risk of hospitalization of vaccinated versus unvaccinated given that an individual is (symptomatically) infected.

The main result in this context is then the following.

**Theorem 2.** Between the quantities $VE_H, VE_H^*$ and $VE$ the relation

$$1 - VE_H = (1 - VE) \cdot (1 - VE_H^*)$$

holds.

**Proof.** Using successively Bayes’ formula, (12) and (1) we obtain

$$VE_H^* = 1 - \frac{P(V | H)}{P(V')} \cdot \frac{P(V')}{P(V)} \cdot \frac{P(I | V')}{P(I | V)}$$

$$= 1 - (1 - VE_H) \cdot \frac{P(I | V')}{P(I | V)}$$

$$= 1 - \frac{1 - VE_H}{1 - VE},$$

and from this the result of the theorem readily follows.

We can therefore compute the real effectiveness regarding hospitalization in terms of the observed effectiveness regarding hospitalization and infection. From Theorem 2 follows

$$1 - VE_H = \frac{1 - VE}{1 - VE_H},$$

a formula which has several consequences. We note that if $VE = 0$, the two measures for effectiveness regarding hospitalization coincide. A simple computation shows that $VE > 0$ implies $VE_H > VE_H^*$, while for $VE < 0$ we have $VE_H < VE_H^*$. We note that in the normal situation of a positive observed effectiveness against symptomatic infections, the real effectiveness regarding hospitalization is smaller than the observed one, in contrast to the relationship between the effectiveness regarding symptomatic infection and the efficacy, which in real life is the other way around. However, in the regime of negative observed effectiveness, typically related to all types of infection, the real effectiveness regarding hospitalization is larger than the observed one.

### 5 Numerical Examples Based on COVID-19 Data

The data from Germany [5], p. 25 and p. 19 respectively, show that in summer 2021, in week 30, the observed effectiveness against symptomatic infections in the age group 18–59 was around $VE = 0.8$ at a vaccination rate of $P(V) = 0.5$, while in November same year, in week 46, the two factors were about $VE = 0.68$ and $P(V) = 0.75$. Assuming a generic plausible value of $VE^* = 0.9$ [10, p. 10], this would imply the values for the degree of discrimination of $D = 0.5$ and $D = 0.69$ in weeks 30 and 46, respectively. This is consistent with the observation that starting with end of August 2021 the so-called 3G-rule was introduced (geimpft-genesen-getestet) which allows access to several public places only with a “green pass” showing the status of vaccinated, recovered or negatively tested. Moreover, since October 11, the fast antigen tests weren’t for free anymore, while starting with Week 46, 3G was replaced with 2G, allowing access only to vaccinated or recovered individuals. The effect of these measures is obviously an increase of the degree of discrimination $D$.

According to the UKHSA [6, p. 33] in weeks 43–46 of 2021, within the age group 40–59 the incidence rates (regarding symptomatic and asymptomatic infections) among the vaccinated individuals were about twice as large as in the case of unvaccinated, implying an observed effectiveness of $VE = -1$. Assuming a degree of discrimination of $D = 0.7$ similar to Germany, this would yield an efficacy of $VE^* = 0.4$ against any infections (not only symptomatic), which is a plausible value. At a degree of discrimination $D = 0.5$ we would have $VE^* = 0$, i.e. an efficacy of 0 against any infections. But such a small value of $D$, similar to the value in Germany in summer 2021, where the cases and also the restrictions were at a low level, is rather unrealistic.

We will consider next the effectiveness against hospitalization and compute the real effectiveness $VE_H^*$ according to (15). In Germany, in week 30 of 2021, for the age group 18–59 we have an observed effectiveness $VE_H = 0.9$ at $VE = 0.8$, so we obtain $VE_H^* = 0.5$. In week 46, for $VE_H = 0.8$ and $VE = 0.68$ we obtain $VE_H^* = 0.375$. Note that the real effectiveness against hospitalization is computed here conditioning on symptomatic infection.

If we compute this quantity by conditioning on any infection, assuming $VE^* = 0.4$ and $D = 0.7$ implies as above the value $VE = -1$. Assuming for example $VE_H = 0.8$ for week 46 in Germany we obtain $VE_H^* = 0.9$. In week 30 we computed $D = 0.5$ and with the assumed value of $VE^* = 0.4$, we obtain for the observed effectiveness conditioned by all infections $VE = -0.2$. Assuming $VE_H = 0.9$, we obtain $VE_H^* = 0.916$.

In the UK, in weeks 43–46, in the age group 40–59 the rates of hospitalization for unvaccinated were about four times larger than for vaccinated, implying $VE_H = 0.75$. As we previously noted, here we have $VE = -1$ considering all infections, which implies $VE_H^* = 0.875$.

The conclusion of this analysis is that the COVID-19-vaccines yield a high real effectiveness of about 90% regarding hospitalization if we condition on any infection, while at conditioning only on symptomatic infections, the
real effectiveness is significantly reduced, at about 40–50% from this value. The interpretation of this fact is that vaccinated individuals develop in most cases only asymptomatic infections, which reduces the hospitalization rate if we condition on all types of infections.

As an additional remark we note that the UK and Germany use different testing strategies. According to the website https://ourworldindata.org/explorers/coronavirus-data-explorer in November 2021 in Germany were performed per 1000 people between 2–3 tests, while in the UK the number is much higher, of about 12–14 tests per 1000 people, so by a factor of 5–6 larger. The main explanation is that the regulations in Germany imply that vaccinated individuals have a reason for testing only if they are symptomatic. Either voluntarily, if one suspects a COVID-19 infection, or compulsory, if the person had contact to a known case and additionally manifests symptoms. Therefore, the asymptomatic infections, especially under the vaccinated, are much underestimated. This explains the fact that in official German statistics the effectiveness of vaccines is evaluated only regarding symptomatic infections. In the UK however, due to the significantly higher test frequency, the estimation of the asymptomatic infections is much better and can be used in the statistics.

6 Summary

In this paper we propose a Bayesian approach linking effectiveness and efficacy of vaccines with the so called degree of discrimination, a factor which quantifies the asymmetric exposure to the virus of vaccinated and unvaccinated individuals. One of the main goals is to explain some statistical data from Germany and the UK in context of the COVID-19 pandemics. Using these data, the theoretical results of this paper lead to the following conclusions on the efficacy of COVID-19 vaccines:

The observed effectiveness of the vaccines depends strongly on the asymmetric higher exposure to the virus of vaccinated individuals as compared with unvaccinated, showing smaller values if the degree of discrimination is high.

While the efficacy regarding symptomatic infections can be high, of over 90%, the efficacy which takes into account also asymptomatic cases is significantly lower, a plausible value being of about 40%. Considering a degree of discrimination larger than this value, this would yield a negative observed effectiveness, as it is the case for the UK.

Considering the effectiveness regarding hospitalization, if we condition on any infection its value is large, of about 90%, while by conditioning only on symptomatic infections, the effectiveness becomes significantly lower, of about 37–50%.

We conclude that the COVID-19 vaccines show in general a high efficacy regarding symptomatic infections and a low efficacy regarding any infections. Vaccinated individuals develop therefore mostly asymptomatic infections, a case in which there is no reason for hospitalization, meaning a high effectiveness regarding hospitalization if conditioning on any infections. However, if conditioned on symptomatic infections, the effectiveness regarding hospitalization is significantly lower.

References


