Bias in vaccine effectiveness studies of clinically severe outcomes that are measured with low specificity: the example of COVID-19-related hospitalisation

Christian Holm Hansen1,2

1. Department of Infectious Disease Epidemiology and Prevention, Statens Serum Institut, Copenhagen, Denmark 2. MRC International Statistics and Epidemiology Group, London School of Hygiene and Tropical Medicine, London, United Kingdom

Correspondence: Christian Holm Hansen (choh@ssi.dk)

Citation style for this article:

Hansen Christian Holm. Bias in vaccine effectiveness studies of clinically severe outcomes that are measured with low specificity: the example of COVID-19-related
hospitalisation. Euro Surveill. 2024;29(7):pii=2300259. htt

Article submitted on 17 May 2023 / accepted on 07 Nov 2023 / published on 15 Feb 2024

Many vaccine effectiveness (VE) analyses of severe disease outcomes such as hospitalisation and death include 'false' cases that are not actually caused by the infection or disease under study. While the inclusion of such false cases inflate outcome rates in both vaccinated and unvaccinated populations, it is less obvious how they affect estimates of VE. Illustrating the main points through simple examples, this article shows how VE is underestimated when false cases are included as outcomes. Depending how the outcome indicator is defined, estimates of VE against severe disease outcomes, whose definition allows for the inclusion of false cases, will be biased downwards and may in certain circumstances approximate the same level as the VE against infection. The bias is particularly pronounced for vaccines that offer high levels of protection against severe disease outcomes but poor protection against infection. Analysing outcomes that are measured with low sensitivity generally does not cause bias in VE studies; defining outcome indicators that minimise the number of false cases rather than the number of missed cases is preferable in VE studies.

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Omicron variant (Phylogenetic Assignment of Named Global Outbreak Lineages (Pangolin) designation B.1.1.529), which appeared towards the end of 2021, was generally associated with milder disease than previous SARS-CoV-2 variants but also with higher transmission rates [1-4]. At the height of the Omicron wave, a large number of hospital patients therefore happened to be SARS-CoV-2-positive despite being hospitalised for reasons unrelated to COVID-19 [5]. Analyses that attributed such hospitalisations to

COVID-19, simply because of the co-occurrence in time with a SARS-CoV-2 infection, exaggerated the number of hospitalisations supposedly due to COVID-19 [6]. But how does such outcome misclassification affect estimates of vaccine effectiveness (VE)?

Vaccine effectiveness is essentially a measure of the protective immunological effect induced by a vaccine against a certain outcome of interest such as infection, disease, hospitalisation or death. It is defined as the ratio of the case rate in vaccinated people to that in unvaccinated people. Valid estimates in VE studies are important as such studies help inform public health decisions. Observational VE studies are routinely used to monitor the protective effect in a population, or segments of a population, of new and established vaccines, for example those against seasonal diseases such as influenza and COVID-19, but they are also used in more unusual outbreaks as seen recently with studies of vaccine protection against mpox [7-9].

In this article, we investigate how outcome misclassification affects estimates in VE studies of severe outcomes. The first section presents the theoretical background and is followed by a section illustrating the main points through six simple scenarios using the example of VE against COVID-19-related hospitalisation and finally some suggestions for assessing the magnitude of the bias based on a few assumptions.

Theoretical background

Define VE against infection with a particular pathogen as $VE_{inf} = 1 - (\pi_{i1}/\pi_{i0})$ where π_{i1} and π_{i0} are the infection rates in the vaccinated and unvaccinated population respectively.

If the outcome of interest is a disease outcome caused by the infection, e.g. hospitalisation, rather than just infection itself, the quantity we seek to estimate is

$$
VE_{hos} = 1 - \tfrac{\pi_{i1} \times \pi_{h1}}{\pi_{i0} \times \pi_{h0}} \ \left(\text{Formula1}\right)
$$

where π_{h_1} and π_{h_0} denote the probability that an infection in a vaccinated and unvaccinated person, respectively, will lead to the outcome. For argument's sake, but without loss of generality, this outcome is taken to be hospitalisation from this point onwards.

In studies using case definitions with low specificity, however, the measured outcome rates will be inflated by the inclusion of misclassified cases. For example, if individuals hospitalised for other reasons also have an 'incidental' infection with the pathogen studied and are erroneously included as cases in the analysis, the quantity actually being estimated is

$$
VE^\prime_{\textit{hos}} = 1 - \tfrac{\pi_{i1} \times \pi_{h1} + \pi_{i1} (1 - \pi_{h1}) \times \pi_f}{\pi_{i0} \times \pi_{h0} + \pi_{i0} (1 - \pi_{h0}) \times \pi_f} \quad \left(\text{Formula 2}\right)
$$

Here π , is the probability that a randomly selected person from the population is hospitalised for reasons unrelated to the studied infection and is assumed independent of vaccination status and infection.

In the numerator of the ratio measure in Formula 2, note that the term $\pi_{i} \times \pi_{i}$ is retained from Formula 1 and represents the proportion of the vaccinated population that is hospitalised due to the infection. The additional term, π_{i_1} (1 – π_{i_2}) × π_f , is the proportion among the vaccinated population that is infected and hospitalised due to other causes. The denominator is of the same form and relates to the unvaccinated population.

We can rewrite Formulae 1 and 2 in terms of relative rates (RR),

$$
\begin{array}{rcl} RR_{hos} \ = \ 1 - VE_{hos} \ = \ \Big(1 - VE_{inf}\Big) \frac{\pi_{h1}}{\pi_{h0}} \ \Big(\text{Formula 3} \Big) \\ RR'_{ \ \ \ \hbox{hos} } \ = \ 1 - VE'_{ \ \ \hbox{hos} } \ = \ \big(1 - VE_{inf}\big) \frac{\pi_{h1} + (1 - \pi_{h1}) \times \pi_f}{\pi_{h0} + (1 - \pi_{h0}) \times \pi_f} \ \Big(\text{Formula 4} \Big) \end{array}
$$

Comparing the expressions in Formulae 3 and 4, and $\text{assuming that } \pi_{h_1} \leq \pi_{h_0}$, it can be seen that $\text{RR}_{hos} \leq \text{RR'}_{hos}$, or equivalently that *VE'hos* ≤*VEhos* , since

$$
\frac{\pi_{h1}}{\pi_{h0}} \leq \frac{\pi_{h1} + (1 - \pi_{h1}) \times \pi_f}{\pi_{h0} + (1 - \pi_{h0}) \times \pi_f}
$$

It can also be seen from Formula 4 that VE'_{hos} tends towards *VE_{inf}* as $π_f$ approaches 1. We therefore have $VE_{hos} ≥ VE²_{hos} ≥ VE³$

By introducing a bias correction factor *c* we can express *RRhos* as a function of *RR'hos*

 $RR_{hos} = c.RR'_{hos}$ (Formula 5)

where
$$
c = \frac{\left(1 - VEin f\right) \frac{\pi_{h1}}{\pi_{h0}}}{\left(1 - VEin f\right) \frac{\pi_{h1} + \left(1 - \pi_{h1}\right) \times \pi_f}{\pi_{h0} + \left(1 - \pi_{h1}\right) \times \pi_f}} = \frac{\pi_{h1}}{\pi_{h0}} \times \frac{\pi_{h0} + \left(1 - \pi_{h0}\right) \times \pi_f}{\pi_{h1} + \left(1 - \pi_{h1}\right) \times \pi_f}
$$

is a value ranging from 1 to π_{h_1}/π_{h_0} as π_f ranges from 0 to 1. Note that the bias is greatest when n_{h_1} is a lot smaller than π_{ho} and when π_f is large relative to π_{ho} and π_{hi} . Meanwhile there is no bias $(c=1)$ when $\pi_{h1} = \pi_{h0}$ since $VE_{has} = VE_{inf}$ at that point. This would be the case for a vaccine which may protect against infection but does not provide any further protection against hospitalisation once infection has occurred. Note also that there is no relationship between the size of the bias, *c,* and the infection rates as expressed through π_{i_0} and π_{i_1} .

The relationship in Formula 5 can be rewritten as $VE_{\text{hos}} = 1 - c(1 - VE_{\text{hos}}^{\prime}).$

Illustrating examples

In the following, the results derived above are demonstrated through a number of scenarios using the example of VE against COVID-19-related hospitalisation. The first three scenarios present examples of a vaccine with a relatively high level of effectiveness against infection, 80%, while the VE against infection in Scenarios $4-6$ is only 20%. The scenarios are hypothetical and have been selected to illustrate the relationship between the VE as estimated in a study and the parameters introduced above.

Scenario 1: Base case scenario

In Scenario 1 (see Table), the proportion of the population who are positive for SARS-CoV-2 is 10% among unvaccinated (π_{io} = 0.10) and 2% among vaccinated people ($π_{i1} = 0.02$), meaning that VE against infection is 80%. Once infected, vaccinated individuals are less likely than unvaccinated individuals to require hospitalisation as only 0.5% of vaccinated cases are hospitalised because of COVID-19 (π_{h_1} =0.005) compared with 2% of unvaccinated cases (π_{ho} =0.02). The true RR of hospitalisation for COVID-19 in the population is therefore (0.02×0.005) / $(0.10 \times 0.02) = 0.05$ resulting in a VE against COVID-19 hospitalisation of 95%. The *estimated* VE of 87.4% is lower, however, due to contamination from individuals with an 'incidental' SARS-CoV-2 diagnosis, i.e. individuals who happen to be infected but who are actually hospitalised for other reasons. Two per cent of the population under study, whether vaccinated or unvaccinated, is in hospital for other reasons $(\pi_{\epsilon} = 0.02)$. The observed COVID-19 hospitalisation rates are consequently inflated in both the vaccinated and unvaccinated group by an amount that is equal to 2% of the total infections observed in the two groups. This forces the numerator and denominator of the RR closer together and pushes down the VE estimate.

Table

SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; VE: vaccine effectiveness.

SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; VE: vaccine effectiveness.
The infection rate is 10% among the unvaccinated population except in Scenario 2 where it is 1%. The risk of hospitalisation due to CO The infection rate is 10% among the unvaccinated population except in Scenario 2 where it is 1%. The risk of hospitalisation due to COVID-19 following an infection is 2% among the unvaccinated population except in Scenario 3 where it is 0.2%. Except in Scenario 6 where it is 0.1%, the risk of hospitalisation in the population among those not hospitalised for reasons related to COVID-19 is 2%, e.g. (20,000– 100) × 0.02=398 in Scenario 1.

Scenario 2: 10 times lower infection rates

As was established above, the incidence rate of infection in the population, reflecting the contagiousness of the pathogen, does not impact the magnitude of this bias (the bias correction factor *c* is independent of π_{i_0} and π_{i_1}). This is illustrated in Scenario 2 where the infection rate is 10 times smaller in both the vaccinated and unvaccinated population $(\pi_{i_0} = 0.01, \pi_{i_1} = 0.002)$ compared with Scenario 1 but the estimated VE remains unchanged.

Scenario 3: 10 times less severe disease

In Scenario 3, the disease is milder as the risk that an infection will lead to hospitalisation is 10 times less that in Scenario 1 (π_{ho} =0.002, π_{hi} =0.0005). Relative to the many misclassified cases with incidental infection, the relevant contrast between vaccinated and unvaccinated cases that are actually hospitalised for COVID-19 is diluted considerably. If the number of hospitalisations with incidental infections is large relative to those actually hospitalised due to the infection under study, the observed hospitalisation rate ratio, which in Scenario 3 is $(400+10) / (1,996+200) = 410/2,196$, will approach the rate ratio for infections among vaccinated vs unvaccinated (20,000/100,000). Consequently, as was established above, we see in Scenario 3 that the estimated VE against hospitalisation approaches the level of VE against infection.

Scenario 4: Poor vaccine effectiveness against infection

When VE for infection is high, as in Scenarios $1-3$, the absolute scale of the bias may be limited. On the other hand, when VE for infection is low (e.g. 20% as in Scenario Δ), estimates of VE against hospitalisation may tend towards similarly low levels despite good vaccine protection against hospitalisation once infected.

Scenario 5: Poor vaccine effectiveness against infection; good vaccine protection once infected

The bias is particularly pronounced for vaccines that offer high levels of protection against hospitalisation despite poor protection against infection as was the case for many of the original (monovalent) COVID-19 vaccines during the Omicron era [10,11]. This is illustrated further in Scenario 5 where the vaccine still only protects 20% against infection but vaccinated SARS-CoV-2 cases are 10 times less likely to require hospitalisation than unvaccinated cases ($π_h$, / π_{ho} =0.002/0.02=0.1). Here, VE against hospitalisation is estimated at just 55.6% when in fact it is 92%.

Scenario 6: Poor vaccine effectiveness against infection; good vaccine protection once infected, low rates of hospitalisations for other reasons

Arguably most important for the magnitude of the bias is the rate of misclassified cases, π , *i.e.* in this example, the proportion of people in hospital for reasons other than COVID-19. This is illustrated in Scenario 6, which is a repeat of Scenario 5 except only 0.1% of the total population under study (not hospitalised for COVID-19) is in hospital for other reasons $(\pi_{\epsilon} = 0.001)$. In this Scenario, the estimated VE of 88.6% is wrong by only a few percentage points. In many real-life scenarios, a lower *π_c* such as in Scenario 6 may be more realistic, especially if the study is conducted in a general population of relatively good health.

Bias correction

As explained above, the true (unbiased) VE against a severe disease outcome, such as hospitalisation, can be expressed as $VE_{\text{hos}}=1-c \times (1-VE_{\text{hoc}}^{\prime})$, where VE_{hos}^{\prime} is the estimated VE and *c* is the bias correction factor

$$
c = \tfrac{\pi_{h1}}{\pi_{h0}} \times \tfrac{\pi_{h0} + (1-\pi_{h0})\pi_f}{\pi_{h1} + (1-\pi_{h1})\pi_f}
$$

which is a function of the three parameters, π_{f} , π_{h1} and π_{h0} . To illustrate, having observed *VE'hos=81.3%* as in Scenario 3, and assumin $g \pi f = 0.02$, $\pi h_1 = 0.0005$ and $\pi h_0 = 0.002$, we can evaluate *c* as 0.2679 and the true VE as $1 - 0.2679 \times (1)$ $-0.813)=95\%$.

In practice $π_ρ$ $π_{h1}$ and $π_{h0}$ will generally not be known but might be gauged from electronic health records or external studies so that it may still be possible to gain a sense of the level of underestimation that can be expected in particular scenarios and, by varying the parameters, suggest a range of plausible values within which the true (unbiased) VE is likely to be.

Discussion

Generally, a low specificity of the outcome measure in clinical studies results in rates being overestimated in both treatment and control groups. Consequently, it is well established that low specificity attenuates the ratio of the measured rates causing the ratio to be closer to 1 than it truly is [12,13]. However, in studies of vaccine protection against severe outcomes, VE is typically derived from the product of two ratios, namely the ratio of infection rates and secondly, among those infected, the ratio of the rates of severe outcome. In many studies, such as those using PCR methods to detect infection, the problem of low specificity affects only the second ratio. Consequently, as illustrated in this article, the biased estimate is bounded by the level of VE against infection.

We have seen that the proportion of individuals admitted to hospital for unrelated reasons, π , is an important factor for the size of the bias. In all the scenarios shown, and in the expression for *c*, it is assumed that π is independent of vaccination status. A fundamental principle of VE studies is that the vaccinated and unvaccinated groups being compared do not differ systematically with respect to other risk factors (at least not after adjustment). This is necessary to ensure that the VE measure captures only the immunological effect of vaccination and not the effects of other exposure and disease predictors. It would probably be a sign that the health profiles are not comparable if π , differed between the two groups, and the resultant VE estimate would therefore also be biased due to confounding. Supplementary analyses with negative control exposures or outcomes are recommended in observational VE studies as a way to assess such bias [14,15]. Nonetheless, it is possible to adapt the expression for c to accommodate different π_c rates, say π_{f_1} and π_{f_0} respectively, in the vaccinated and unvaccinated group; it can then be shown that the underestimation is even more pronounced if π_{f_1} \rightarrow π_{f_0} , which would be the situation if, for example, people of poorer health were more likely to be vaccinated.

Observational studies to estimate VE may be designed in various ways. Study designs include retrospective cohort studies through analysis of routinely collected electronic health records, cross-sectional designs, prospective cohort studies, test-negative designs and other case–control studies. The biasing effects presented in this article, caused by outcome misclassification, apply equally across all these designs. Even randomised controlled trials (RCTs) are vulnerable to this type of outcome misclassification bias, although due to their budgets and rigour, RCTs often include more accurate outcome assessment methods than observational studies.

Unlike the problems caused by low specificity, low sensitivity of an outcome measure generally does not bias the ratio of the measured rates in the two groups [13]. It is therefore preferable for studies of VE to use outcomes that minimise numbers of misclassified cases (false cases) rather than numbers of missed cases.

Depending on data availability, minimising the number of misclassified cases may be achieved by defining more specific outcome measures based for example on primary diagnosis codes upon hospital admission, hospital procedures or death certificate information. Death as a specific disease outcome may also be defined with higher specificity, but possibly at the cost of lower sensitivity, by requiring certain accompanying hospital diagnoses or procedures. A number of studies have explored alternative severe outcome definitions in the context of VE [6,16,17].

Whether in the context of COVID-19, influenza or some other type of infection, studies of VE against severe disease outcomes such as hospitalisation and death should aim to use outcome measures that minimise inclusion of false cases to avoid underestimation of the effects of interest. Where this is not possible, consideration should be given to the magnitude of the resulting bias, for example by investigating likely scenarios for the three parameters that enter the expression for *c* above.

Conclusions

Vaccine effectiveness against severe disease outcomes will generally be underestimated when incidental cases are included in the analysis. The potential error is greatest for vaccines that offer high levels of protection against severe disease but poor protection against infection. Outcomes with high specificity rather than high sensitivity are preferable in VE studies.

Ethical statement

Ethical approval was not needed for this work because no participant data were involved.

Funding statement

CHH is partly funded through a UK Medical Research Council (MRC) grant (reference: MR/R010161/1). The funder had no role in the study, decision to publish, or preparation of the manuscript.

Conflict of interest

None declared.

Authors' contributions

CHH alone conceived the idea and is responsible for the manuscript and decision to publish.

References

- World Health Organization (WHO). Classification of omicron (B.1.1.529): SARS-CoV-2 variant of concern. Geneva: WHO: 2021. Available from: https://www.who.int/news/item/26-11- 2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variantof-concern
- 2. Bálint G, Vörös-Horváth B, Széchenyi A. Omicron: increased transmissibility and decreased pathogenicity. Signal Transduct Target Ther. 2022;7(1):151. https://doi.org/10.1038/s41392- 022-01009-8 PMID: 35525870
- Bager P, Wohlfahrt J, Bhatt S, Stegger M, Legarth R, Møller CH, et al. , Omicron-Delta study group. Risk of hospitalisation associated with infection with SARS-CoV-2 omicron variant versus delta variant in Denmark: an observational cohort study. Lancet Infect Dis. 2022;22(7):967-76. https://doi. org/10.1016/S1473-3099(22)00154-2 PMID: 35468331
- 4. Nyberg T, Ferguson NM, Nash SG, Webster HH, Flaxman S, Andrews N, et al. , COVID-19 Genomics UK (COG-UK) consortium. Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: a cohort study. Lancet. 2022;399(10332):1303-12. https://doi. org/10.1016/S0140-6736(22)00462-7 PMID: 35305296
- 5. Voor In 't Holt AF, Haanappel CP, Rahamat-Langendoen J, Molenkamp R, van Nood E, van den Toorn LM, et al. Admissions to a large tertiary care hospital and Omicron BA.1 and BA.2 SARS-CoV-2 polymerase chain reaction positivity: primary, contributing, or incidental COVID-19. Int J Infect Dis. 2022;122(2):665-8. https://doi.org/10.1016/j.ijid.2022.07.030 PMID: 35842214
- 6. Stowe J, Andrews N, Kirsebom F, Ramsay M, Bernal JL. Effectiveness of COVID-19 vaccines against Omicron and Delta hospitalisation, a test negative case-control study. Nat Commun. 2022;13(1):5736. https://doi.org/10.1038/s41467- 022-33378-7 PMID: 36180428
- 7. Price AM, Flannery B, Talbot HK, Grijalva CG, Wernli KJ, Phillips CH, et al. Influenza vaccine effectiveness against influenza A(H3N2)-related illness in the United States during the 2021- 2022 influenza season. Clin Infect Dis. 2023;76(8):1358-63. https://doi.org/10.1093/cid/ciac941 PMID: 36504336
- 8. Poland GA, Kennedy RB, Tosh PK. Prevention of monkeypox with vaccines: a rapid review. Lancet Infect

Dis. 2022;22(12):e349-58. https://doi.org/10.1016/S1473- 3099(22)00574-6 PMID: 36116460

- 9. Deputy NP, Deckert J, Chard AN, Sandberg N, Moulia DL, Barkley E, et al. Vaccine effectiveness of JYNNEOS against mpox disease in the United States. N Engl J Med. 2023;388(26):2434-43. https://doi.org/10.1056/ NEJMoa2215201 PMID: 37199451
- 10. Gram MA, Emborg H-D, Schelde AB, Friis NU, Nielsen KF, Moustsen-Helms IR, et al. Vaccine effectiveness against SARS-CoV-2 infection or COVID-19 hospitalization with the Alpha, Delta, or Omicron SARS-CoV-2 variant: A nationwide Danish cohort study. PLoS Med. 2022;19(9):e1003992. https://doi. org/10.1371/journal.pmed.1003992 PMID: 36048766
- 11. Andrews N, Stowe J, Kirsebom F, Toffa S, Rickeard T, Gallagher E, et al. Covid-19 vaccine effectiveness against the Omicron (B.1.1.529) variant. N Engl J Med. 2022;386(16):1532-46. https://doi.org/10.1056/NEJMoa2119451 PMID: 35249272
- 12. Jackson ML, Rothman KJ. Effects of imperfect test sensitivity and specificity on observational studies of influenza vaccine effectiveness. Vaccine. 2015;33(11):1313-6. https://doi. org/10.1016/j.vaccine.2015.01.069 PMID: 25659280
- 13. Field trials of health interventions: a toolbox. Smith PG, Morrow RH, Ross DA. 3rd edition. Oxford: Oxford University Press; 2015. pp. 211-212. PMID: 26225404
- 14. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 2010;21(3):383-8. https://doi. org/10.1097/EDE.0b013e3181d61eeb PMID: 20335814
- 15. Hansen CH, Moustsen-Helms IR, Rasmussen M, Søborg B, Ullum H, Valentiner-Branth P. Short-term effectiveness of the XBB.1.5 updated COVID-19 vaccine against hospitalisation in Denmark: a national cohort study. Lancet Infect Dis. 2024;24(2):e73-4. https://doi.org/10.1016/S1473- 3099(23)00746-6 PMID: 38190834
- 16. Chung H, Austin PC, Brown KA, Buchan SA, Fell DB, Fong C, et al. Effectiveness of COVID-19 vaccines over time prior to omicron emergence in Ontario, Canada: test-negative design study. Open Forum Infect Dis. 2022;9(9):ofac449. https://doi. org/10.1093/ofid/ofac449 PMID: 36147595
- 17. World Health Organization (WHO). Evaluation of COVID-19 vaccine effectiveness: Interim guidance. Geneva: WHO; 2021. Available from: https://www.who.int/publications/i/item/ WHO-2019-nCoV-vaccine_effectiveness-measurement-2021.1

License, supplementary material and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence and indicate if changes were made.

Any supplementary material referenced in the article can be found in the online version.

This article is copyright of the authors or their affiliated institutions, 2024.