Modeling bias in vaccine trials relying on fragmented healthcare records

Peter J. Yim, PhD[1]*

Abstract
COVID-19 vaccine trials depend on the localization of vaccination records for each trial subject. Misclassification bias occurs when vaccination records cannot be localized or uniquely identified. This bias may be significant in trials where the trial subjects' vaccination and health records are distributed between more than one database. The potential for this bias is present in numerous published COVID-19 vaccine trials. A model is proposed for estimation of the magnitude of this bias on apparent vaccine efficacy. In the model, misclassification is always in the direction from partial or fully vaccinated status to unvaccinated status. The model predicts a disproportionate effect of vaccination status misclassification on the apparent vaccine efficacy when population vaccination rates are high.

Keywords
Vaccine efficacy, Bias, Misclassification

[1] *Virtual Scalpel, Inc., Belle Mead, NJ, USA*
*__Corresponding author__: yimpjp@virtualscalpel.com

1. Introduction

Randomized trials of COVID-19 vaccines available in the United States showed efficacy for reducing the risk of symptomatic disease [1], [2] and severe disease [3]. However, the application of the vaccines to a broader segment of the population, against an evolving disease and with respect to different outcomes including mortality has required additional study.

Observational trials have reported high levels of efficacy of the COVID-19 vaccines against severe outcomes [4], [5], [6], [7], [8], [9] [10], [11], [12], [13] but also an efficacy that wanes [14], [15]. These studies may be playing an important role in forming public perceptions of vaccine efficacy. Thus, these observational studies should be carefully evaluated.

The data collection methodologies in the observational studies are complex and have a direct bearing on the outcomes of the studies. In all the observational studies cited above, records of vaccinations and healthcare outcomes are maintained on separate information systems, for at least some of the study subjects. In all but the study of Green *et al* [10], comparison of COVID-19 health outcomes with vaccination status requires matching of subject records between information systems. When the matching is incomplete, subjects are assigned the "unvaccinated" status, by default, thus introducing a bias.

This bias has not been assessed in any of the cited studies; the upper bound on this bias is not known. The effect could be minor or it could be invalidating. As a result, the results of those studies should be considered uncertain.

The purpose of this study is to present a model relating misclassification to the apparent vaccine efficacy and to use the model to better understand this bias.

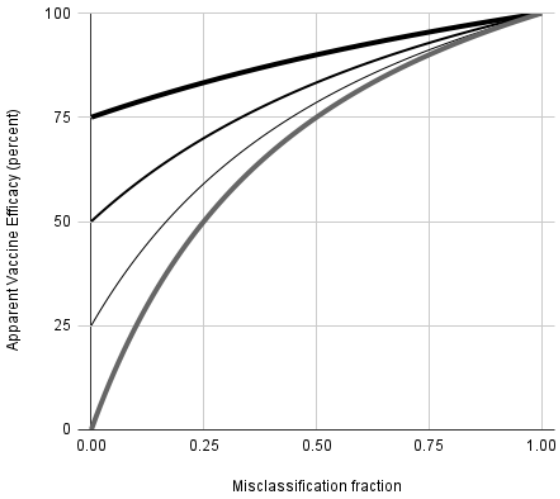

Figure 1. Example of effect of vaccination misclassification on apparent vaccine efficacy. .In this case, the unvaccinated, partially vaccinated and fully vaccinated populations are equal. Also, the relative risk of the partially and fully vaccinated are equal.

Symbol	Definition
VE	Vaccine efficacy
VEA	Apparent vaccine efficacy
N_U	Unvaccinated population
N_P	Partially vacccinated population
N_F	Fully vaccinated population
E_U	Unvaccinated events
E_P	Partially vaccinated events
E_F	Fully vaccinated events
R_U	Unvaccinated risk
R_P	Partially vaccinated risk
R_F	Fully vaccinated risk
RR_U	Relative risk unvaccinated
RR_P	Relative risk partially vaccinated
RR_F	Relative risk fully vaccinated
m	Misclassification fraction

Table 1. Table of Symbols

2. Background

The magnitude of this bias in the COVID-19 vaccine trials is unknown. However, some minimum misclassification rates have been reported for vaccination status. In the study of Lin *et al*, non-inclusion of vaccination records from federal entities resulted an estimated 5% rate of missing data [4]. In the study of Olson *et al* [5], vaccination records could not be verified for 4% of hospitalized patients based on the hospital patient record and approximate vaccination date and a "reasonable" location.

2.1 Non-compliance with reporting requirements

There are diverse reasons why vaccination records may be unavailable. In California, non-compliance of vaccination providers for submission of vaccination records affected the completeness of the state registry [16]. The study of Tartof *et al* [8] did rely on the California immunization registry so it is possible that non-compliance of vaccination providers did affect the outcome.

2.2 Non-reporting due to consent requirement

Completeness of the public record of vaccination record can also be limited by consent requirements. The NYC Health Department's Citywide Immunization Registry, for example requires consent from all adults for submission of their vaccination record [17]. That registry was used for the vaccine trial of Rosenberg *et al* [11].

2.3 Incomplete matching due to data-entry variation

Missing vaccination records can also result from variation in data entry between vaccination registries and other databases. In the study of Lin *et al* [4], a probabilistic algorithm was used to match records from the North Carolina Covid-19 Surveillance System and the Covid-19 Vaccine Management System. Also, as alluded to in Lin *et al*, vaccination status may be recorded in either or both the state and federal registries. In the trial by the Washington State Department of Health [7], vaccination status was established only if exact matches on the first name, last name and date of birth between the Washington Immunization Information System and Washington Disease Reporting System.

3. Methods

A model is proposed for the effect of vaccination status misclassification on apparent vaccine efficacy. The model assumes only one way misclassification from partial and full vaccination status to unvaccinated status due to missing data.

The model applies to static conditions. In this model, the proportion of unvaccinated, partially vaccinated and fully vaccinated populations are constant. Furthermore, the relative risks associated with partial and full vaccination relative to non-vaccination are constant.

Vaccine efficacy VE is defined in terms of the risk of a COVID-19 outcome for the unvaccinated population R_U and the fully vaccinated population R_F:

$$VE = \frac{R_U - R_F}{R_U} \tag{1}$$

In terms of the COVID-19 events E_U and E_F in the unvaccinated and fully vaccinated populations, respectively and in terms of the unvaccinated and fully vaccinated populations, N_U and N_F, respectively:

$$VE = \frac{\frac{E_U}{N_U} - \frac{E_F}{N_F}}{\frac{E_U}{N_U}} \tag{2}$$

The fraction of events that are misclassified as unvaccinated is m. The apparent vaccine efficacy VEA is obtained by the addition of misclassified events from the partially vaccinated population, $m(E_P)$ and of the fully vaccinated population $m(E_F)$ to the events in the unvaccinated population. Also, the misclassified events are reduced accordingly in the fully vaccinated population:

$$VEA = \frac{(\frac{E_U + m(E_P) + m(E_F)}{N_U}) - (\frac{E_F - m(E_F)}{N_F})}{\frac{E_U + m(E_P) + m(E_F)}{N_U}} \tag{3}$$

The apparent vaccine efficacy can be expressed in terms of the risks in the unvaccinated, partially vaccinated and fully vaccinated populations, R_U, R_P and R_F, respectively. N_P is the population with partial vaccination.

$$VEA = \frac{R_U + m(\frac{N_P}{N_U})(R_P) + m(\frac{N_F}{N_U})(R_F) - R_F + m(R_F)}{R_U + m(\frac{N_P}{N_U})(R_P) + m(\frac{N_F}{N_U})(R_F)} \tag{4}$$

	Vaccine Efficacy	Outcome	Time frame
Lin *et al* [4]	94.1% *	Hospitalizaton/Mortality	6 months post-vaccination
Olson *et al* [5]	98% **	ICU admission	July 1 - October 25, 2021
Johnson *et al* [6]	94% ††	Mortality	July - November, 2021
Department of Health, Washington State [7]	88.8% ††	Mortality	December 15,2021 - January 11, 2022
Tartof *et al* [8].	93%**	Hospitalization	Through 6 months post-vaccination
Bruxvoort *et al* [9]	95.8%*	Hospitalization and in-hospital mortality	Through June 30, 2021
Greene *et al* [10]	Not determined	Hospitalization and Mortality	February 21–April 17, 2021
Rosenberg *et al* [11]	92.6%††	Hospitalization	Week of June 21, 2021
Puranik *et al* [12]	91.6%*	Hospitalization	January - July, 2021
Thompson *et al* [13]	90% †	ICU Admission	January 1 through June 22, 2021

Table 2. Vaccine efficacy against severe outcomes in studies of COVID-19 relying on fragmented public health records. If more than one type of outcome reported, results for the most severe outcome shown. If results for more than one vaccine reported, the best case outcome shown.*mRNA-1273, *BNT162b2, †mRNA-1273 and BNT162b2, ††all vaccinations

Finally, the apparent vaccine efficacy can be expressed in terms of the relative risks ratios in the partially vaccinated and fully vaccinated populations, RR_P and RR_F, respectively.

$$VEA = \frac{1 + m(\frac{N_P}{N_U})(RR_P) + m(\frac{N_F}{N_U})(RR_F) - RR_F + m(RR_F)}{1 + m(\frac{N_P}{N_U})(RR_P) + m(\frac{N_F}{N_U})(RR_F)} \tag{5}$$

4. Results

The relationship between the misclassification fraction m and the apparent vaccine efficacy VEA was examined for the case where $N_U = N_P = N_F$ and for the cases of $VE = 0\%$, $VE = 25\%$, $VE = 50\%$ and $VE = 75\%$. The results are shown in figure 1.

5. Discussion

The effect of vaccination status misclassification has been inadequately addressed in COVID-19 vaccine efficacy trials relying on fragmented public health records. Thus, there is potential that the reported vaccine efficacy from these trials may be significantly overstated.

References

[1] Thomas SJ, Moreira ED, Kitchin N, Absalon J, Gurtman A, Lockhart S, Perez JL, Pérez Marc G, Polack FP, Zerbini C, Bailey R, Swanson KA, Xu X, Roychoudhury S, Koury K, Bouguermouh S, Kalina WV, Cooper D, Frenck RW, Hammitt LL, Türeci Ö, Nell H, Schaefer A, Ünal S, Yang Q, Liberator P, Tresnan DB, Mather S, Dormitzer PR, Şahin U, Gruber WC, and Jansen KU. Safety and efficacy of the bnt162b2 mRNA Covid-19 vaccine through 6 months. *New England Journal of Medicine*, 385(19):1761–1773, 2021.

[2] El Sahly HM, Baden LR, Essink B, Doblecki-Lewis S, Martin JM, Anderson EJ, Campbell TB, Clark J, Jackson LA, Fichtenbaum CJ, Zervos M, Rankin B, Eder F, Feldman G, Kennelly C, Han-Conrad L, Levin M, Neuzil KM, Corey L, Gilbert P, Janes H, Follmann D, Marovich M, Polakowski L, Mascola JR, Ledgerwood JE, Graham BS, August A, Clouting H, Deng W, Han S, Leav B, Manzo D, Pajon R, Schödel F, Tomassini JE, Zhou H, and Miller J. Efficacy of the mRNA-1273 SARS-CoV-2 vaccine at completion of blinded phase. *New England Journal of Medicine*, 385(19):1774–1785, 2021.

[3] Sadoff J, Gray G, Vandebosch A, Cárdenas V, Shukarev G, Grinsztejn B, Goepfert PA, Truyers C, Fennema H, Spiessens B, Offergeld K, Scheper G, Taylor KL, Robb Merlin L, Treanor J, Barouch DH, Stoddard J, Ryser MF, Marovich MA, Neuzil KM, Corey L, Cauwenberghs N, Tanner T, Hardt K, Ruiz-Guiñazú J, Le Gars M, Schuitemaker H, Van Hoof J, Struyf F, and Douoguih M. Safety and efficacy of single-dose Ad26.COV2.S vaccine against Covid-19. *New England Journal of Medicine*, 384(23):2187–2201, 2021.

[4] Lin D-Y, Gu Y, Wheeler B, Young H, Holloway S, Sunny S-K, Moore Z, and Zeng D. Effectiveness of covid-19 vaccines over a 9-month period in North Carolina. *New England Journal of Medicine*, 0(0):null, 2022.

[5] Olson SM, Newhams MM, Halasa NB, Price AM, Boom JA, Sahni LC, Pannaraj PS, Irby K, Walker TC, Schwartz SP, Maddux AB, Mack EH, Bradford TT, Schuster JE, Nofziger RA, Cameron MA, Chiotos K, Cullimore ML, Gertz SJ, Levy ER, Kong M, Cvijanovich NJ, Staat

Study	Classification method/criteria	Misclassification
Lin *et al* [4]	Probabilistic match, Link Plus 3.0	At least 5%
Olson *et al* [5]	Study personnel	At least 4%
Johnson *et al* [6]	”active linkage”	Not specified
Department of Health, Washington State [7]	Exact match	”some”
Tartof *et al* [8].	Not specified	Not specified
Bruxvoort *et al* [9]	Vaccination data imported into healthcare system database manually and electronically	”is possible”
Greene *et al* [10].	Not required	NA
Rosenberg *et al* [11]	”exact algorithms”	Not specified
Puranik *et al* [12]	”automated biweekly syncing”	Not specified
Thompson *et al* [13]	Not specified	”misclassification of vaccine exposures or outcomes could bias our vaccine-effectiveness estimates downward”

Table 3. Strategies for matching vaccination and healthcare outcomes and comments on efficacy in observational vaccine trials.

MA, Kamidani S, Chatani BM, Bhumbra SS, Bline KE, Gaspers MG, Hobbs CV, Heidemann SM, Maamari M, Flori HR, Hume JR, Zinter MS, Michelson KN, Zambrano LD, Campbell AP, Patel MM, and Randolph AG. Effectiveness of BNT162b2 vaccine against critical Covid-19 in adolescents. *New England Journal of Medicine*, 0(0):null, 2022.

[6] Johnson AG, Amin AB, and et al Ali AR. Covid-19 incidence and death rates among unvaccinated and fully vaccinated adults with and without booster doses during periods of delta and omicron variant emergence — 25 U.S. jurisdictions, April 4–December 25, 2021. *MMWR Morb Mortal Wkly Rep 2022;71:132–138.*.

[7] Washington State Department of Health. Covid-19 cases, hospitalizations, and deaths by vaccination status. February 2, 2022.

[8] Tartof SY, Slezak JM, Fischer H, Hong V, Ackerson BK, Ranasinghe ON, Frankland TB, Ogun OA, Zamparo JM, Gray S, Valluri SR, Pan K, Angulo FJ, Jodar L, and McLaughlin JM. Effectiveness of mrna bnt162b2 covid-19 vaccine up to 6 months in a large integrated health system in the usa: a retrospective cohort study. *Lancet.*, 398(10309):1407–1416, 2021.

[9] Bruxvoort KJ, Sy LS, Qian L, Ackerson BK, Luo Y, Lee GS, Tian Y, Florea A, Takhar HS, Tubert JE, Talarico CA, and Tseng HF. Real-world effectiveness of the mrna-1273 vaccine against Covid-19: Interim results from a prospective observational cohort study. *Lancet Reg Health Am.*, 2022.

[10] Greene SK, Levin-Rector A, McGibbon E, Baumgartner J, Devinney K, Ternier A, Sell J, Kahn R, and Kishore N. Reduced COVID-19 hospitalizations among New York City residents following age-based SARS-CoV-2 vaccine eligibility: Evidence from a regression discontinuity design. *Vaccine: X*, 10(100134), 2022.

[11] Rosenberg ES, Holtgrave DR, Dorabawila V, Conroy M, Greene D, Lutterloh E, Backenson B, Hoefer D, Morne J, Bauer U, and Zucker HA. New COVID-19 cases and hospitalizations among adults, by vaccination status - New York, May 3-July 25, 2021. *MMWR Morb Mortal Wkly*, 78(37):1306–1311, 2021.

[12] Puranik A, Lenehan PJ, Silvert E, Niesen N, Corchado-Garcia J, O'Horo JC, Virk A, Swift MD, Halamka J, Badley AD, Venkatakrishnan AJ, and Soundararajan V. Comparison of two highly-effective mRNA vaccines for COVID-19 during periods of Alpha and Delta variant prevalence. *medRxiv*, 2021.

[13] Thompson MG, Stenehjem E, Grannis S, Ball SW, Naleway AL, Ong TC, DeSilva MB, Natarajan K, Bozio CH, Lewis N, Dascomb K, Dixon BE, Birch RJ, Irving SA, Rao S, Kharbanda E, Han J, Reynolds S, Goddard K, Grisel N, Fadel WF, Levy ME, Ferdinands J, Fireman B, Arndorfer J, Valvi NR, Rowley EA, Patel P, Zerbo O, Griggs EP, Porter RM, Demarco M, Blanton L, Steffens A, Zhuang Y, Olson N, Barron M, Shifflett P, Schrag SJ, Verani JR, Fry A, Gaglani M, Azziz-Baumgartner E, and Klein NP. Effectiveness of Covid-19 vaccines in ambulatory and inpatient care settings. *New England Journal of Medicine*, 385(15):1355–1371, 2021.

[14] Goldberg Y, Mandel M, Bar-On YM, Bodenheimer O, Freedman L, Haas EJ, Milo R, Alroy-Preis S, Ash N, and Huppert A. Waning immunity after the bnt162b2 vaccine

in Israel. *New England Journal of Medicine*, 385(24):e85, 2021.

[15] Chemaitelly H, Tang P, Hasan MR, AlMukdad S, Yassine HM, Benslimane FM, Al Khatib HA, Coyle P, Ayoub HH, Al Kanaani Z, Al Kuwari E, Jeremijenko A, Kaleeckal AH, Latif AN, Shaik RM, Abdul Rahim HF, Nasrallah GK, Al Kuwari MG, Al Romaihi HE, Butt AA, Al-Thani MH, Al Khal A, Bertollini R, and Abu-Raddad LJ. Waning of BNT162b2 vaccine protection against SARS-CoV-2 infection in Qatar. *New England Journal of Medicine*, 385(24):e83, 2021.

[16] Money L. California's digital COVID-19 vaccination record has glitches. Here's how to fix yours. *Los Angeles Times*, June 21, 2021.

[17] https://www1.nyc.gov/site/doh/services/cir-parents-guardians.page.