# Is the Observed Protection of COVID-19 Vaccines Against Infection within 14 days Real or an Artifact? A Negative Control Outcomes-Based Investigation Using Real-World Data

**Bingyu Zhang, MS[1,2^], Qiong Wu, PhD[1,3,4^], Ting Zhou, MD, PhD[1,3], Dazheng Zhang, MS[1,3], Jiayi Tong, PhD[1,3,5], Yuqing Lei, MS[1,3], Martijn J. Schuemie, PhD[6,7,8], Patrick B. Ryan, PhD[6,9], Jeffrey S. Morris, PhD[3], George Hripcsak, MD, MS[10], Christopher B. Forrest, MD, PhD[11], Yong Chen, PhD[1,2,3]**

**[1] The Center for Health AI and Synthesis of Evidence (CHASE), University of Pennsylvania, Philadelphia, PA, USA**
**[2] Applied Mathematics and Computational Science, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA**
**[3] Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA**
**[4] Department of Biostatistics and Health Data Science, University of Pittsburgh, Pittsburgh, PA, USA**
**[5] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA**
**[6] Observational Health Data Sciences and Informatics, New York, NY, USA**
**[7] Epidemiology, Office of the Chief Medical Officer, Johnson & Johnson, Titusville, NJ, USA**
**[8] Department of Biostatistics, University of California, Los Angeles, CA, USA**
**[9] Global Epidemiology Organization, Johnson & Johnson, Titusville, NJ, USA**
**[10] Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, USA**
**[11] Applied Clinical Research Center, Children's Hospital of Philadelphia, Philadelphia, PA, USA**

^: Co-first authors

## Background

Observational studies and real-world data (RWD) have made significant contributions to medical research and public health. Successful examples include LEGEND studies[1], Nurses' Health Study (NHS)[2], and statins for coronary heart disease (CHD)[3]. However, generating reliable real-world evidence (RWE) from RWD is challenging due to unmeasured confounders and data quality issues, known as residual bias. Negative control outcome (NCO) experiments[4,5], which use clinical outcomes believed to have no causal relationship to the exposure, have become an important tool for addressing these biases; see examples such as Suchard et al.[1], Wu et al.[6], among many.

During the COVID-19 pandemic, RWD has played a pivotal role in advancing our understanding and informing policy decisions to combat the virus. Comparative effectiveness research (CER) and target trial emulation (TTE) studies on vaccine effectiveness have been particularly critical[6–12]. However, these studies have reported varying results regarding the effectiveness of COVID-19 vaccines, with effectiveness estimates ranging widely across studies.

An important ongoing debate in this field is whether infections occurring within the first week(s) after vaccination should be considered as negative controls[9,10]. Dagan et al.[9] observed a consistent pattern of

similarity between the comparison groups during the first 12 days after the first dose, which served as a "negative control" period.

On the other hand, Ostropolets et al.[10] argued that the truth might be the opposite. They found unexpectedly high effectiveness in the first week post-vaccination, and through chart reviews they suggested that vaccinated patients are more likely to attribute their symptoms to common vaccine side effects, making them less likely to seek medical care.

In this paper, we aim to investigate this key question, i.e., whether SARS-CoV-2 infection within 14 days after vaccination should be considered a negative control outcome of COVID-19 vaccines, which is instrumental in subsequent comparative effectiveness studies using observational data for vaccine studies. To address this issue, we conducted a formal hypothesis testing procedure[4] to investigate whether SARS-CoV-2 infection within 14 days after vaccination is a negative control outcome of COVID-19 vaccines, where the null hypothesis is that the infection is truly a negative control outcome.

We then test this hypothesis using the EHR data from the PEDSnet[13], a large-scale pediatric electronic health record (EHR) cohort. By carefully examining the initial period after vaccination, we aim to enhance the understanding of vaccine performance and address the discrepancies observed in previous studies.

## Methods

### DATA SOURCES

The data were contributed by 8 members of PEDSnet[13], a nationwide learning health collaboration of pediatric health systems covering 7 million patients, including Children's Hospital of Philadelphia, Cincinnati Children's Hospital Medical Center, Children's Hospital Colorado, Ann & Robert H. Lurie Children's Hospital of Chicago, Nationwide Children's Hospital, Nemours Children's Hospital, Nemours Children's Hospital, Seattle Children's Hospital, and Stanford Children's Health.

### STUDY DESIGN

We conducted a target trial emulation from January 1, 2022, to November 16, 2022. The study included patients aged 12 to 20 at cohort entry, who had no previously documented SARS-CoV-2 infection, no prior COVID-19 vaccination, had at least one visit 24 months before the index date, and had no visits before the index date.

The intervention of interest was the first dose of any COVID-19 vaccine, in comparison with no receipt of any type of COVID-19 vaccine. Vaccine information was identified based on the immunization domain, using either the presence of a CVX code designating an administered or patient-reported dose or of a source value containing the terms "COVID" or "SARS" in the immunization table.

During the study period, all newly vaccinated individuals were matched in a 1:1 ratio to unvaccinated controls on each day. Newly vaccinated individuals were eligible for inclusion even if they had previously served as controls, and an unvaccinated comparator could serve as a match for vaccinated persons in more than one sequential trial as long as they remained unvaccinated. For example, persons vaccinated on day 1 were matched with unvaccinated persons on day 1; persons vaccinated on day 2 were matched with individuals who remained unvaccinated from day 1 to 2, and so forth. Unvaccinated participants were assigned an index date corresponding to the day they entered the trial. Follow-up for each participant ended at the earliest of the following events: documented SARS-CoV-2 infection, vaccination (for unvaccinated controls), vaccination of the matched control (for vaccinated persons), or the end of the

study period. We performed exact matching on the following patient characteristics: age (in years), gender (female, male), race/ethnicity (NHW, Hispanic, NHB, Multiple, Other/Unknown), indicators from the eight data-contributing sites, and Pediatric Medical Complexity Algorithm[14] (PMCA, no chronic condition, non-complex chronic condition, complex chronic condition).

The outcome of interest was documented SARS-CoV-2 infection within 14 days following cohort entry. A documented SARS-CoV-2 infection was defined by positive polymerase-chain-reaction (PCR), serology, antigen tests, or diagnoses of COVID-19, or diagnoses of post-acute sequelae of SARS-CoV-2 (PASC) or multisystem inflammatory syndrome in children associated with COVID-19 (MIS-C). The outcome date was set as either the earliest date of positive tests, COVID-19 diagnoses, PASC diagnoses, or MIS-C diagnoses.

**Figure 1** summarizes the participant selection for the vaccinated and unvaccinated groups, and **Figure 2** shows the study design for target trial emulation assessing vaccine effectiveness.

STATISTICAL ANALYSIS

The absolute number and cumulative incidence of documented SARS-CoV-2 infection in the vaccinated and unvaccinated groups were reported and estimated with the Kaplan-Meier estimator[15], during days 1 through 14 after the first dose of vaccine (**Figure 3**).

We used the modified Poisson regression model for binary outcomes to estimate the risk ratio (RR) for documented SARS-CoV-2 infection between vaccinated and unvaccinated groups while adjusting for different follow-up lengths among participants, and 95% confidence intervals (CIs) were estimated with robust sandwich variance estimators. The vaccine effectiveness (VE) was estimated as (1-RR) *100%.

We conducted two secondary analyses to understand the potential health-seeking behavior and time-varying effects of the COVID-19 vaccine. Since COVID-19 vaccines have no biological effect on flu infections, the first analysis focused on documented flu infections within 14 days following cohort entry. This allowed us to examine infection patterns where health-seeking behavior is likely similar to that of patients with documented SARS-CoV-2 infection. The second analysis extended beyond the 14-day window to evaluate whether the estimated vaccine effectiveness against SARS-CoV-2 infection, as well as any observed trends in flu infections, varied over time.

Furthermore, we performed an additional sensitivity analysis using Cox regression, Poisson regression, and logistic regression to estimate hazard ratios (HRs), relative risks (RRs), and odds ratios (ORs) between the comparison groups for both periods of the first 7 and 14 days. Analyses were performed using R version 4.4.0.

NCO-BASED INVESTIGATIONS

To assess to validity of candidate NCOs, we performed a hypothesis test comparing the estimated effect of the outcome of interest to the empirical distribution of pre-specified NCOs. Suppose the estimated effect size for the outcome of interest is $\hat{\eta}$ with standard error $\hat{\tau}$. For the general NCOs, we estimate the empirical mean $\hat{\mu}$ and standard error $\hat{\sigma}$.

The null hypothesis states that the candidate outcome is exchangeable with the general NCOs, implying that it is a valid NCO. We computed a calibrated two-sided p-value based on the empirical null distribution[4]:

$$2 * \min\left\{\Phi\left(\frac{\hat{\eta}-\hat{\mu}}{\sqrt{\hat{\sigma}^2+\hat{\tau}^2}}\right), 1-\Phi\left(\frac{\hat{\eta}-\hat{\mu}}{\sqrt{\hat{\sigma}^2+\hat{\tau}^2}}\right)\right\},$$

where $\Phi$ is the cumulative distribution function of the standard normal distribution. A p-value <0.05 indicates statistical significance, meaning the null hypothesis is rejected, and the candidate NCO is not valid. In such cases, the candidate NCO is not exchangeable with the general NCOs, suggesting it may be influenced by the exposure.

**Results**

COHORT IDENTIFICATION

Within the PEDSnet database, we identified 42,641 adolescents aged 12 to 20 years who received a first dose COVID-19 vaccine before November 16, 2022. A total of 25,555 individuals were eligible for the study and 25,146 were matched to unvaccinated controls (**Figure 1**). Among these patients, 23,130 (46.0%) were male with a median (Q1, Q3) age of 14 (13, 16) years. Data for 9.9% of the unvaccinated controls and their matched pairs were censored when the controls received the vaccine.

VACCINE EFFECTIVENESS WITHIN 14 DAYS

During the follow-up of the first 14 days after cohort entry, 43 and 153 infections were documented in vaccinated and unvaccinated groups, respectively. **Figure 3** shows the cumulative incidence curves for the documented SARS-CoV-2 infection. A moderate protective VE of 71.2% was observed (95% CI, 60.6 to 80.0) for documented SARS-CoV-2 infection.

Additionally, a moderate protective VE of 69.2% (95% CI, 53.0 to 79.9) was observed within the first 7 days after vaccination. The results were consistent across different modeling approaches, with the Cox model showing a VE of 72.0% (95% CI, 60.7 to 80.0) and the logistic model showing a VE of 60.0% (95% CI, 60.7 to 80.0) for documented SARS-CoV-2 infection within the first 14 days after cohort entry. More detailed results are provided in **Table 1**.

NCO-BASED INVESTIGATION

**Figure 4** shows each general NCO's estimated log relative risk, alongside documented SARS-CoV-2 and flu infections, within and beyond 14 days following cohort entry. The dark green line indicates the mean of the empirical distribution of general NCOs, with the shaded bands representing the 95% confidence intervals. **Figure 5** displays the EASE plot for general NCOs and documented SARS-CoV-2 and flu infections, within and beyond 14 days after COVID-19 vaccination.

Within the 14-day period, the documented SARS-CoV-2 infections deviated from the distribution of general NCOs, which implies it is not a valid NCO for COVID-19 vaccination.

FLU-BASED INVESTIGATION

For documented flu infection within the first 14 days after cohort entry, a high protective VE of 88.2% (95% CI, 49.1 to 97.3). Beyond 14 days post-vaccination, the estimated VE was 40.7% (95% CI, 29.2 to 50.4) for documented flu infection. The null hypothesis cannot be rejected (P=0.151).

Within the 14-day period, documented flu infections deviated from the distribution of general NCOs, which indicates that it is not a valid NCO for COVID-19 vaccination. In contrast, beyond 14 days, the distribution of documented flu infections was indistinguishable from that of other general NCOs.

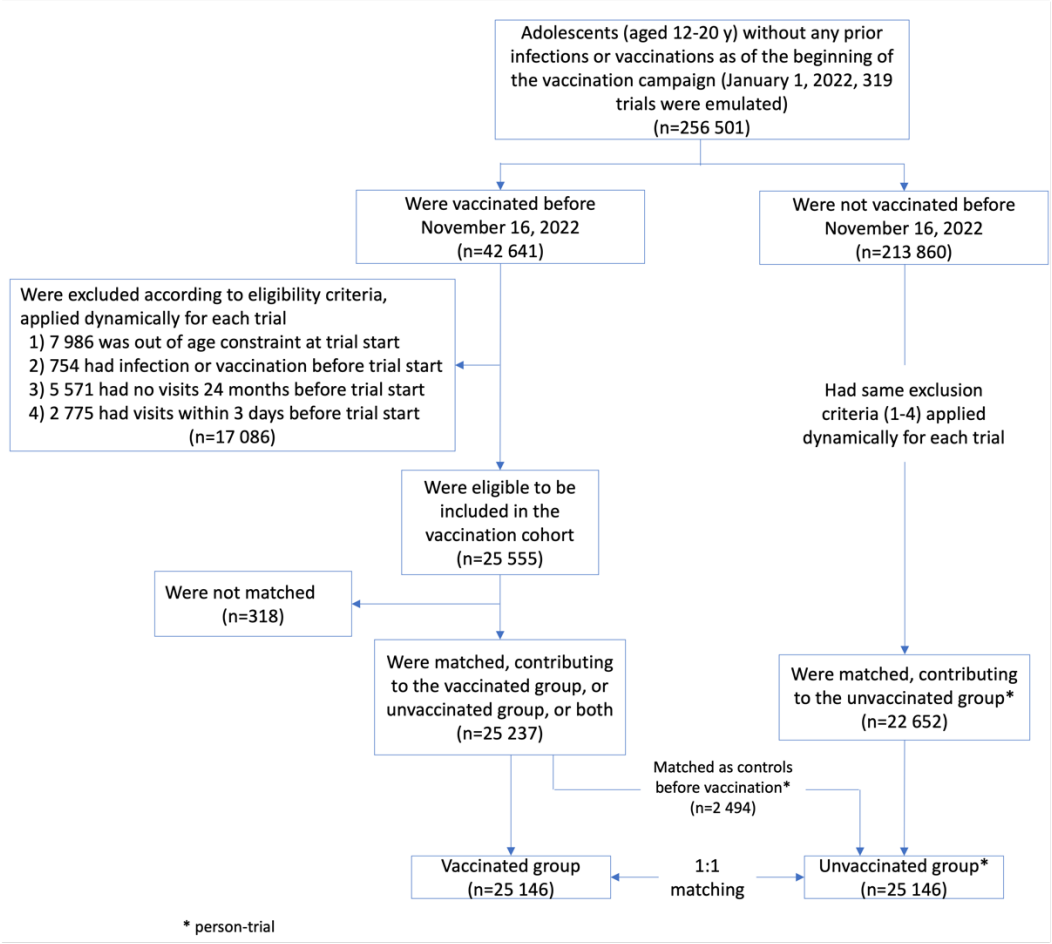


**Figure 1. Study population and cohort enrollment process.**

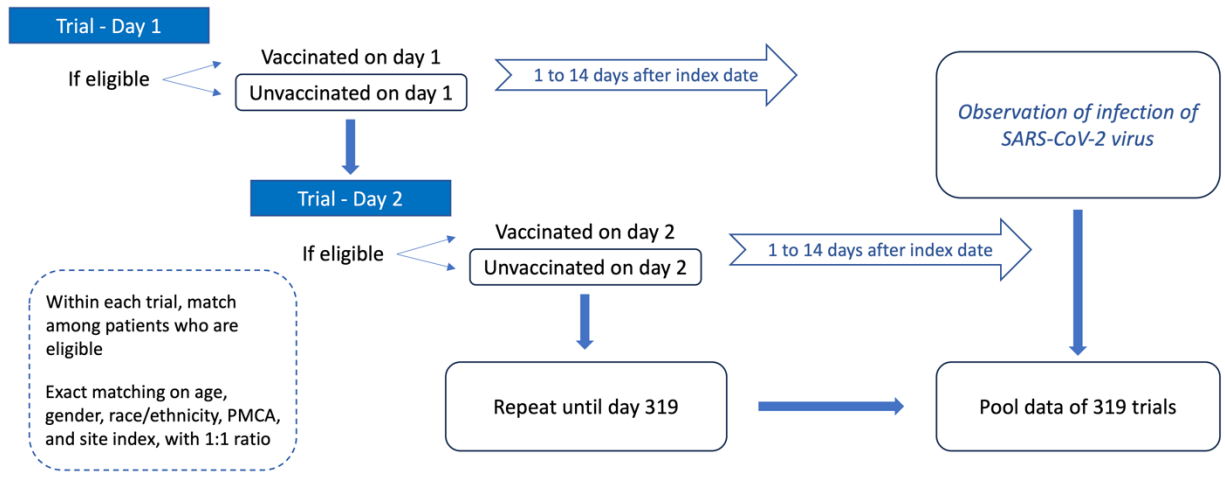


**Figure 2. Study Design for target trial emulation assessing vaccine effectiveness by comparing the vaccinated group to the unvaccinated group in terms of documented SARS-CoV-2 infection within 14 days after the index date.**

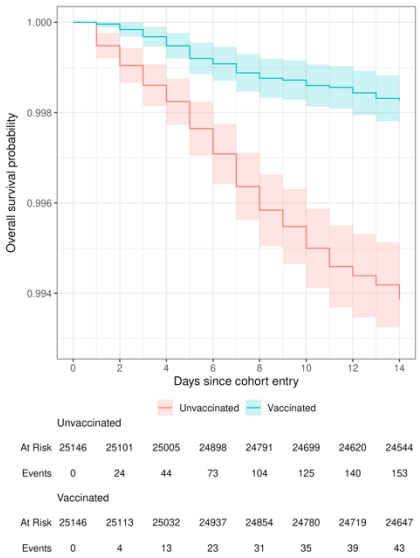


**Figure 3. Cumulative incidence of documented SARS-CoV-2 infections in vaccinated and unvaccinated groups.**

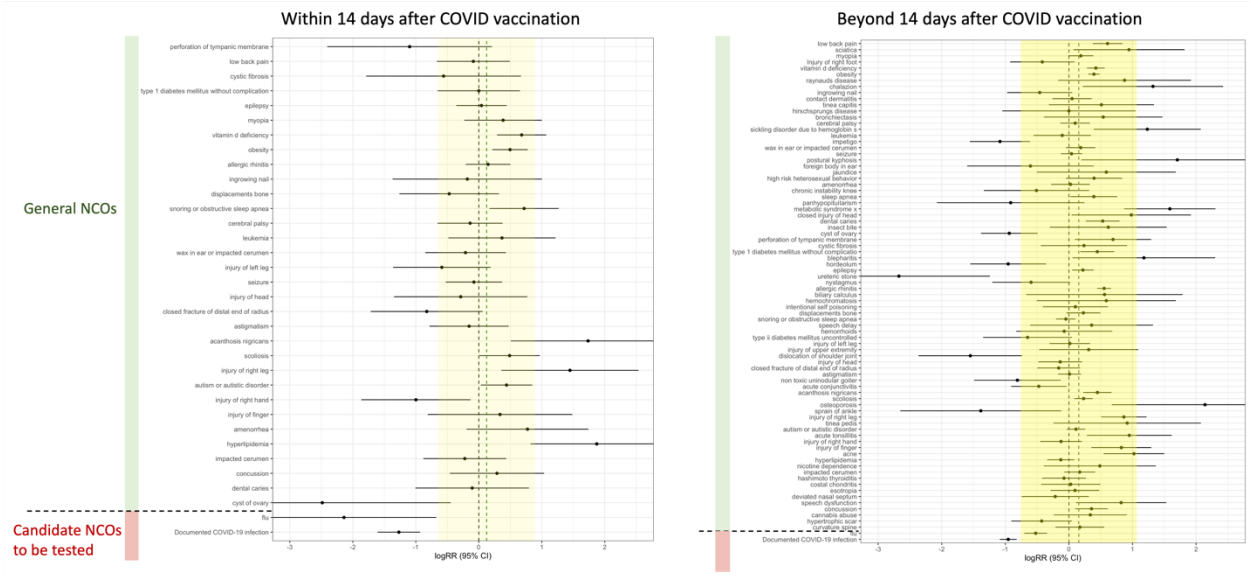


**Figure 4. Forest plot of log relative risk of general negative control outcomes, documented SARS-CoV-2 infection, and documented flu infection within and beyond 14 days after COVID-19 vaccination.**

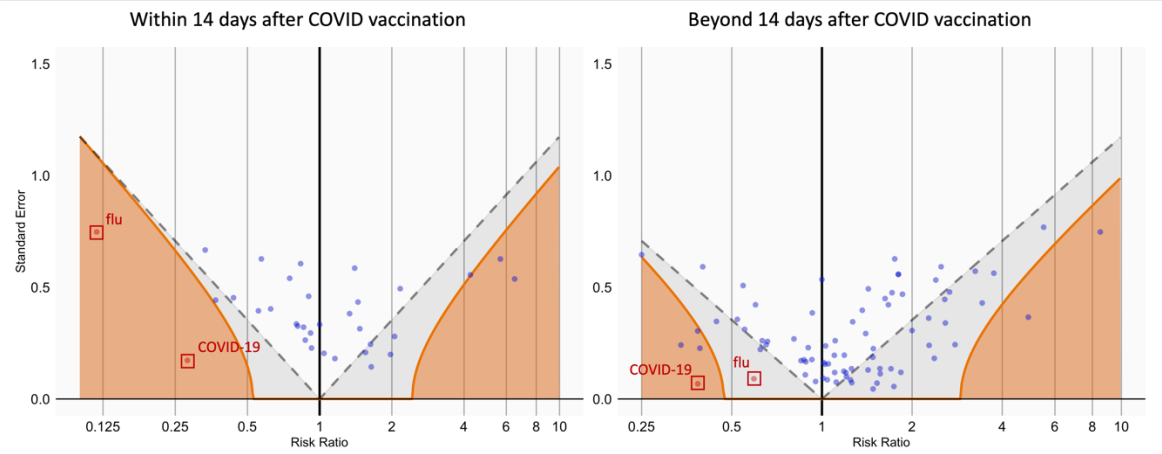


**Figure 5. EASE plot and counter-enhanced funnel plot of general negative control outcomes, documented SARS-CoV-2 infection, and documented flu infection within and beyond 14 days after COVID-19 vaccination.**

**Table 1. Estimated mean and standard error of the risk of infection within the initial period after vaccination, and the estimated mean and standard error of the empirical null distribution for Cox, Poisson, and logistic regression models across different initial days.**

| Model | Days | Log risk ratio of COVID-19 infection | Standard error of log risk ratio of COVID-19 infection | Log risk ratio general NCOs | Standard error of log risk ratio of general NCOs | P value |
|---|---|---|---|---|---|---|
| **Cox** | 7 | -1.180 | 0.216 | 0.107 | 0.363 | <0.001 |
| **Cox** | 14 | -1.271 | 0.173 | 0.108 | 0.360 | <0.001 |
| **Poisson** | 7 | -1.179 | 0.216 | 0.300 | 0.001 | <0.001 |
| **Poisson** | 14 | -1.269 | 0.173 | 0.255 | 0.235 | <0.001 |
| **Logistic** | 7 | -1.181 | 0.216 | 0.300 | 0.000 | <0.001 |
| **Logistic** | 14 | -1.274 | 0.173 | 0.240 | 0.240 | <0.001 |

## Conclusion

In this study, we utilized a sequential trial emulation design on a large-scale pediatric EHR cohort from PEDSnet to examine the characterization of documented SARS-CoV-2 infection following COVID-19 vaccination during the Omicron era. Our results indicate that documented SARS-CoV-2 infection within the initial week(s) after vaccination differs significantly from the commonly used NCOs in the OHDSI community. This finding holds true across different time intervals examined (7- and 14-days post-vaccination) and across various outcome models employed (Cox proportional hazards, Poisson regression, and logistic regression). In addition, documented flu infections within 14 days also deviated from the expected distribution of general NCOs, which implies it cannot serve as a valid NCO despite the lack of a direct biological relationship with the COVID-19 vaccine. Beyond 14 days, the distribution of documented flu infections was indistinguishable from that of other general NCOs.

A plausible explanation, as suggested by Ostropolets et al.[10], is that the observed protective effect of vaccine incidence may be confounded by vaccine-related changes in health-seeking behavior. Specifically, following vaccination, individuals experiencing flu or COVID-like symptoms may attribute these symptoms

to the vaccine's side effects rather than a possible SARS-CoV-2 infection. Consequently, they may be less likely to seek medical care and instead choose to stay at home, leading to an underreporting of COVID-19 infection cases in the immediate post-vaccination period.

These results highlight that the early period after vaccination should be carefully handled in observational studies assessing vaccine effectiveness.

## References


1. Suchard MA, Schuemie MJ, Krumholz HM, et al. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *The Lancet* 2019; 394: 1816–1826.
2. Bhupathiraju SN, Grodstein F, Rosner BA, et al. Hormone therapy use and risk of chronic disease in the nurses' health study: a comparative analysis with the Women's Health Initiative. *Am J Epidemiol* 2017; 186: 696–708.
3. Danaei G, Rodríguez LAG, Cantero OF, et al. Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. *Stat Methods Med Res* 2013; 22: 70–96.
4. Schuemie MJ, Ryan PB, DuMouchel W, et al. Interpreting observational studies: why empirical calibration is needed to correct p-values. *Stat Med* 2014; 33: 209–218.
5. Schuemie MJ, Hripcsak G, Ryan PB, et al. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proceedings of the National Academy of Sciences* 2018; 115: 2571–2577.
6. Wu Q, Tong J, Zhang B, et al. Real-World Effectiveness of BNT162b2 Against Infection and Severe Diseases in Children and Adolescents. *Ann Intern Med*.
7. Chemaitelly H, AlMukdad S, Ayoub HH, et al. Covid-19 vaccine protection among children and adolescents in Qatar. *New England Journal of Medicine* 2022; 387: 1865–1876.
8. Cohen-Stavi CJ, Magen O, Barda N, et al. BNT162b2 vaccine effectiveness against omicron in children 5 to 11 years of age. *New England Journal of Medicine* 2022; 387: 227–236.
9. Dagan N, Barda N, Kepten E, et al. BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. *New England Journal of Medicine* 2021; 384: 1412–1423.
10. Ostropolets A, Hripcsak G. COVID-19 vaccination effectiveness rates by week and sources of bias: a retrospective cohort study. *BMJ Open* 2022; 12: e061126.
11. Pilishvili T, Gierke R, Fleming-Dutra KE, et al. Effectiveness of mRNA Covid-19 vaccine among US health care personnel. *New England Journal of Medicine* 2021; 385: e90.
12. Hall VJ, Foulkes S, Saei A, et al. COVID-19 vaccine coverage in health-care workers in England and effectiveness of BNT162b2 mRNA vaccine against infection (SIREN): a prospective, multicentre, cohort study. *The Lancet* 2021; 397: 1725–1735.
13. Forrest CB, Margolis PA, Bailey LC, et al. PEDSnet: a national pediatric learning health system. *Journal of the American Medical Informatics Association* 2014; 21: 602–606.
14. Simon TD, Haaland W, Hawley K, et al. Development and validation of the Pediatric Medical Complexity Algorithm (PMCA) version 3.0. *Acad Pediatr* 2018; 18: 577–580.
15. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958; 53: 457–481.