medRxiv preprint doi: https://doi.org/10.1101/2022.11.30.22282923; this version posted November 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

November 30, 2022

The imprinting effect of COVID-19 vaccines: an expected selection bias in observational studies

Short title: Expected selection bias in imprinting studies

Susana Monge[1,2], Roberto Pastor-Barriuso[1,3], Miguel A. Hernán[4]

[1] National Centre of Epidemiology, Institute of Health Carlos III, Madrid, Spain.

[2] Consortium for Biomedical Research in Infectious Diseases (CIBERINFEC), Spain

[3] Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Spain

[4] CAUSALab and Departments of Epidemiology and Biostatistics, Harvard T H Chan School of Public Health, Boston, MA, USA

Abstract

Findings of recent observational studies have been interpreted as supporting immune imprinting of COVID-19 vaccines. In this work, we clarify that the current discussion can be mapped to an attempt to estimate the direct effect of vaccine boosters on SARS-CoV-2 reinfections, and that such direct effect cannot be correctly estimated with observational data. We conclude that recent observational estimates regarding immune imprinting are fundamentally biased, and that the increased risk of reinfection in individuals vaccinated with a vaccine booster compared to no booster is expected even if the immune imprinting hypothesis is false. We use graphical methods (directed acyclic graphs), data simulations and analysis of real-life data to illustrate the mechanism and magnitude of this bias.

Keywords:

Selection bias; collider bias; causality; SARS-CoV-2; imprinting; vaccines

PRE-PRINT

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.11.30.22282923; this version posted November 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

The SARS-CoV-2 Omicron variant and subvariants have significant antigenic changes compared with both previous variants and COVID-19 vaccines used until September 2022. There is concern that past exposure to previous variants—through infection or vaccination—can alter the immunological response to an Omicron infection in such a way that the immune response to successive Omicron infections would be impaired (1,2). Under this so-called immune imprinting, receiving a booster (third dose) might increase the risk of re-infection with Omicron.

If a pernicious effect of immune imprinting truly existed, current recommendations for additional vaccine doses may need to be re-evaluated. The findings of recent observational studies have been interpreted as supporting this immune imprinting hypothesis. The analysis of observational data shows indeed an increased risk of Omicron re-infection in individuals vaccinated with three doses of Wuhan-based monovalent vaccines compared with two doses (3) (and no increased risk of Omicron re-infection in unvaccinated individuals (4)). This interpretation, however, is misguided because such increased risk is expected even if the immune imprinting hypothesis is false.

In this work, we clarify that the current discussion about immune imprinting can be mapped to an attempt to estimate the direct effect of vaccine boosters, and that such direct effect cannot be correctly estimated with observational data. We conclude that recent observational estimates regarding immune imprinting (3,4) are fundamentally biased. First, we show how the causal question of interest can be precisely articulated.

Specification of the target trial

The immune imprinting hypothesis states that a vaccine booster in individuals who later are infected by Omicron increases the risk of a second Omicron infection. A useful procedure to precisely articulate a causal question is to describe the hypothetical randomized experiment—the target trial (5)—that would answer it. The target trial may be impractical or unethical, but that is beside the point.

Consider a target trial in which the elegibility criteria are ≥18 years of age, having received the second dose of an mRNA vaccine at least 90 days ago and having received no third dose yet, no previous laboratory-confirmed SARS-CoV-2 infection, and not being part of a population with special vaccination recommendations (e.g., no nursing home residents, institutionalized individuals, or health-care workers). The intervention would have two components. First, eligible persons would be randomly assigned to 1) immediate administration of a booster (third

medRxiv preprint doi: https://doi.org/10.1101/2022.11.30.22282923; this version posted November 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

dose) of an mRNA COVID-19 vaccine, or 2) no further vaccine doses. Second, all participants would be forced to remain uninfected until we infect them with Omicron at a random time within, say, 6 months of randomization. The outcome of interest would be a laboratory-confirmed Omicron infection at least 90 days after the first Omicron infection.

Individuals would be followed from assignment. However, because the outcome cannot occur (by definition) until 90 days after completing the intervention, the cumulative incidence curves for both groups would stay at zero until 90 days after randomization. Therefore, under the assumption that no reinfection can truly occur during the first 90 days, individuals will be followed from 90 days after the randomized first infection until the earliest of Omicron reinfection (outcome of interest), death, or administrative end of follow-up (9 months after randomization).

Using the data from this target trial, we could quantify the direct effect of a booster on the risk of reinfection by comparing the risk of reinfection between individuals assigned to booster and no booster. We refer to this causal effect as the (controlled) direct effect because it quantifies the effect of the booster that is not mediated through the first Omicron infection.

Emulation of the target trial is not possible

The target trial described above, however, is unfeasible because, in the real world, we cannot force people to get infected at a time of our choice. When a target trial cannot be carried out, we often use observational data from human populations to emulate it. In this case, however, it is not possible to use observational data to emulate the target trial without bias, as we will explain.

Some observational studies have tried to estimate the direct effect of the booster on reinfection by 1) adjusting for factors that may confound the effect of the booster on reinfection, and 2) restricting the analysis to individuals who had their first infection after the booster (3). Let us suppose that confounding adjustment 1) was successful and therefore the observational study appropriately accounts for the lack of randomization of the booster. Even in that setting, the restriction 2) on having had a first Omicron infection is expected to introduce selection bias (6) because, in the real world, the first infection occurs more frequently among people with higher susceptibility. Therefore, if the booster prevents infections, it is essentially guaranteed that persons who received the booster and subsequently had a first infection are, on average, more susceptible to reinfection that persons who did not receive the booster and subsequently had a first infection. In the absence of data on individual susceptibility, an observational study cannot unbiasedly estimate the direct effect of a booster.

PRE - PRINT

medRxiv preprint doi: https://doi.org/10.1101/2022.11.30.22282923; this version posted November 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

To see this graphically, consider the simplified causal directed acyclic graphs (DAGs) in Figure 1. The first causal DAG represents the (randomized) target trial and the second causal DAG represents the observational analysis that restricts to persons with a first infection. The causal DAGs include the nodes booster (yes, no), confirmed infection in period 1 (yes, no), confirmed infection in period 2 (yes, no), and an unmeasured "susceptibility" variable that represents individual characteristics that increase the risk of infection (e.g., subclinical immunosuppression, occupational and behavioral factors) or of receiving a diagnosis of infection (e.g., testing behavior, access to the health system).

In the target trial (Figure 1a), there is no arrow from booster to infection in period 1 which, by design, occurs in all individuals, regardless of whether they were assigned to booster or no booster. Therefore, the unconditional association between booster and infection in period 2 is an unbiased estimator of the direct effect of booster on reinfection (the doted arrow) if everybody had had a first infection. Figure 1a also represents a variation of the target trial in which individuals would have been randomly assigned to infection in period 1; in that trial there would be no arrow from booster to first infection either.

In the analysis of observational data restricted to people with infection in period 1 (Figure 1b), there is an arrow from booster to infection in period 1 because the booster reduce the risk of infection (7,8), and there is a box around infection in period 1 to represent the conditioning of the analysis on individuals with value "yes" on that variable. In graph theory, we say that infection in period 1 is a collider. Therefore, restricting to those with infection in period 1 equal to "yes" is form of collider stratification, which is expected to induce a noncausal association between booster and reinfection (9). That is, the association between booster and reinfection among those with a first infection combines the direct effect of booster on reinfection (the doted arrow) and the selection bias induced by the open path through susceptibility (6).

If the booster had no direct effect (i.e., if the dotted arrow did not exist) then the risk of infection in period 2 would be expected to be greater for those who did vs. did not receive the booster. This higher risk in the booster group vs. the no booster group is the entire result of selection bias and thus has no causal interpretation as a harmful effect of the booster on reinfection. In fact, all this elevated risk indicates is that people who get infected despite receiving a booster are people more susceptible to reinfection.

We designed a simplified simulation to quantify the magnitude of the selection bias under the causal DAG in Figure 1b. We simulated a dataset of 10 million persons with a normally distributed susceptibility variable, of whom 65% were randomly assigned to booster. We

medRxiv preprint doi: https://doi.org/10.1101/2022.11.30.22282923; this version posted November 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

assumed that the booster decreased the probability of infection in period 1 by 50% and had no effect on the probability of infection in period 2, i.e., there was no direct effect of booster on reinfection. For a realistic risk of infection of 10% in period 1, the (non-causal) odds ratio of infection in period 2 for booster vs. no booster ranged between 1.04 and 1.37, depending on the assumed distribution of susceptibility in the population (see the Supplement for details and computer code). As expected, restricting the observational analysis to individuals with a first Omicron infection results in a higher risk of reinfection in the booster group even if the booster had zero effect on reinfection. It is all selection bias.

Replication of the selection bias in real world data

We attempted to emulate the target trial described above using nationwide observational data from Spain. To do so, we linked individual-level data from three Spanish population registries (Vaccination Registry [REGVACU], Laboratory Results Registry [SERLAB], and National Health System [NHS] registry), as described elsewhere (8).

We used the observational data to identify individuals eligible for the target trial, starting on January 1, 2022. We assigned those who received a booster to the booster group and, for each of them, we randomly chose a matched individual who did not receive a booster on the same week. The matching factors included sex, age (±5 years), province, time since primary vaccination (±14 days) and type of primary vaccination (BNT162b2 or mRNA-1273). In an attempt to emulate a target trial in which all participants are infected with Omicron within 10 months of booster assignment, we restricted the analysis to individuals with a laboratory-confirmed SARS-CoV-2 infection in the next 10 months (and further matched eligible individuals on week of infection). That is, we only considered for the analysis individuals with an Omicron infection and further matched them 1:1 on week of infection. As discussed in the previous section, restricting to individuals with infection induces uncontrollable selection bias because of differential selection between groups that depends on individual susceptibility.

We then followed individuals starting on day 90 after infection until the earliest of a confirmed SARS-CoV-2 reinfection, death, discontinuation of registration in the NHS database, or administrative censoring (October 31, 2022). To estimate the per-protocol effect, we censored at receipt of any additional vaccine dose. The cumulative incidence (risk) in each group was estimated using the Kaplan Meier method (10) and compared between groups via risk ratios (RR). Nonparametric bootstrapping with 500 samples was used to compute percentile-based 95% confidence intervals (95%CI).

PRE-PRINT

medRxiv preprint doi: https://doi.org/10.1101/2022.11.30.22282923; this version posted November 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Of 12,749,506 initially eligible individuals, 1,704,904 experienced a first Omicron infection in the study period; of them, 425,741 (25%) had received a booster dose before the infection. We could exactly match 249,226 (59%) individuals with a booster to the same number of controls, with a median age of 44 years. A total of 201,266 (81%) matched pairs remained under follow-up 90 days after the infection and were included in the analysis. During a maximum follow-up of 211 days (mean 133) there were 1,794 re-infections, with a 6-month risk over the full period of 0.59% in the booster group and 0.54% in the control group. The risk ratio of reinfection in the booster group compared with the no booster group was 1.08 (0.97, 1.20) at 6 months of follow-up (9 months post-infection); but varied between 1.03 (95%CI: 0.93, 1.17) in days 0 to 90 of follow-up and 1.20 (95%CI: 0.98-1.45) in days 91 to 180.

Discussion

We have used graphical methods and simulations to explain that recent observational estimates of an apparently greater risk of Omicron reinfection after a booster dose cannot be endowed with a causal interpretation. Rather, observational estimates restricted to individuals with an earlier Omicron infection can be fully explained by selection bias. We illustrated the selection bias by conducting a real world analysis of nationwide data from Spain. Our estimates of increased risk of reinfection in individuals infected after receiving a booster was compatible with our simulation results and comparable with those from previous observational studies (3).

Removing the selection bias would require the measurement, and adjustment for, individual susceptibility to infection or diagnosis. Unfortunately, this information is not available. Comorbidities or health seeking-behavior are unlikely to fully capture individual susceptibility and, in fact, studies accounting for some of these measured factors (3) did not provide different estimates to the one in our study that accounted only for age, sex, location, and type of vaccine.

Of course, several other possible sources of bias may exist in observational studies of vaccine effectiveness, including cofounding from incomplete adjustment for prognostic factors associated with vaccination, and measurement error from incomplete ascertainment of SARS-CoV-2 infection. Here we focused on the selection bias that is expected to arise in any studies that condition on post-vaccination infection, including randomized trials that, like essentially all trials, cannot intervene on infection itself.

In summary, analyses of observational data require a precise articulation of the causal question before the estimates can be interpreted. Observational analyses to estimate the direct effect of a booster on the risk of reinfection (i.e., “imprinting”) failed to specify the target trial that they were trying to emulate. As a result, an elevated risk of reinfection among individuals who

PRE-PRINT

medRxiv preprint doi: https://doi.org/10.1101/2022.11.30.22282923; this version posted November 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

received a booster and had a first post-booster infection was incorrectly interpreted as demonstrating a harmful effect of the booster. An explicitly causal approach to these questions indicates that 1) the elevated risk is mathematically expected and may be fully explained by selection bias and 2) observational data may not be generally used to answer these "imprinting" questions.

Ethical statement

The use of the NHS database, REGVACU and SERLB for the purpose of monitoring vaccine effectiveness has been approved by the research ethics committee at the Instituto de Salud Carlos III (CEI PI 98_2020 and CEI PI 08_2022). Informed consent was not required because this study is based on national population registries.

Data availability statement

The databases used in the observational data analysis are owned by the Ministry of Health and the Autonomous Communities in Spain, which establish the requirements for their access and use.

Patient and Public Involvement statement

It was not appropriate or possible to involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research.

Acknowledgements

We acknowledge the contribution of everyone that makes possible to have real-time data on COVID-19 vaccination and laboratory tests available in Spain, including professionals in the 19 Autonomous Communities and Cities, the Vaccines Division and Health Information Systems Department of the Ministry of Health, and the National Centre of Epidemiology at the Institute of Health Carlos III.

PRE - PRINT

medRxiv preprint doi: https://doi.org/10.1101/2022.11.30.22282923; this version posted November 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Contributions

MAH and SM conceived the study and simulations, RP and SM performed the simulations, SM performed the analyses. SM is the guarantor of this article. The corresponding author attests that all listed authors meet the authorship criteria and that no others meeting the criteria have been omitted.

Funding

There was no specific funding provided for the study

Conflicts of interest

Authors declare no conflicts of interest. MH is data science adviser for ProPublica and consultant for Cytel.

References

1. Reynolds CJ, Pade C, Gibbons JM, Otter AD, Lin KM, Muñoz Sandoval D, et al. Immune boosting by B.1.1.529 (Omicron) depends on previous SARS-CoV-2 exposure. Science. 2022 Jul 15;377(6603):eabq1841.

2. Röltgen K, Nielsen SCA, Silva O, Younes SF, Zaslavsky M, Costales C, et al. Immune imprinting, breadth of variant recognition, and germinal center response in human SARS-CoV-2 infection and vaccination. Cell. 2022 Mar 17;185(6):1025-1040.e14.

3. Chemaitelly H, Ayoub HH, Tang P, Coyle P, Yassine HM, Al Thani AA, et al. COVID-19 primary series and booster vaccination and immune imprinting. medRxiv. 2022 Jan 1;2022.10.31.22281756.

4. Chemaitelly H, Ayoub HH, Tang P, Hasan MR, Coyle P, Yassine HM, et al. Immune Imprinting and Protection against Repeat Reinfection with SARS-CoV-2. N Engl J Med. 2022 Nov 3;387(18):1716–8.

5. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016 Apr 15;183(8):758–64.

6. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiol Camb Mass. 2004 Sep;15(5):615–25.

7. Abu-Raddad LJ, Chemaitelly H, Ayoub HH, AlMukdad S, Yassine HM, Al-Khatib HA, et al. Effect of mRNA Vaccine Boosters against SARS-CoV-2 Omicron Infection in Qatar. N Engl J Med. 2022 Mar 9;

8. Monge S, Rojas-Benedicto A, Olmedo C, Mazagatos C, José Sierra M, Limia A, et al. Effectiveness of mRNA vaccine boosters against infection with the SARS-CoV-2 omicron

PRE-PRINT

medRxiv preprint doi: https://doi.org/10.1101/2022.11.30.22282923; this version posted November 30, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

(B.1.1.529) variant in Spain: a nationwide cohort study. Lancet Infect Dis. 2022 Jun 2;S1473-3099(22)00292-4.

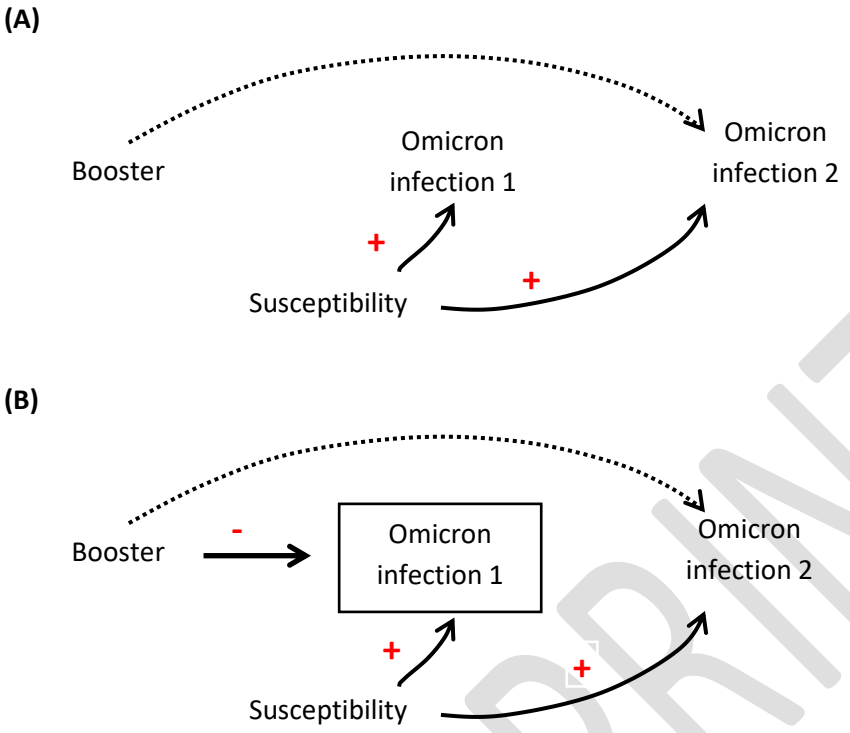
9. Pearl J (2009). Causality. Cambridge University Press.

10. Kaplan EL MP. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 53:457–81.

PRE - PRINT

medRxiv preprint doi: https://doi.org/10.1101/2022.11.30.22282923; this version posted November 30, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1. Causal directed acyclic graphs representing (A) a hypothetical target trial and (B) a naïve analysis of observational data.



PRE - PRINT