The extent and impact of vaccine status miscategorisation on covid-19 vaccine efficacy studies

Martin Neil (m.neil@qmul.ac.uk)
School of Electronic Engineering and Computer Science
Queen Mary, University of London

Norman Fenton (n.fenton@qmul.ac.uk)
School of Electronic Engineering and Computer Science
Queen Mary, University of London

Scott McLachlan (scott.mclachlan@kcl.ac.uk)
School of Nursing, Midwifery and Palliative Care
Kings College London

Abstract

It is recognised that many studies reporting high efficacy for Covid-19 vaccines suffer from various selection biases. Systematic review identified thirty-eight studies that suffered from one particular and serious form of bias called miscategorisation bias, whereby study participants who have been vaccinated are categorised as unvaccinated up to and until some arbitrarily defined time after vaccination occurred. Simulation demonstrates that this miscategorisation bias artificially boosts vaccine efficacy and infection rates even when a vaccine has zero or negative efficacy. Furthermore, simulation demonstrates that repeated boosters, given every few months, are needed to maintain this misleading impression of efficacy. Given this, any claims of Covid-19 vaccine efficacy based on these studies are likely to be a statistical illusion.

Keywords: simulation; covid-19; evidence-based medicine; mis-categorisation; selection bias; observational studies; public health; vaccine effectiveness.

1. Introduction

Considerable attention has been given to the reported high efficacy for the Covid-19 vaccines and how many of these studies have exhibited signs of selection bias (Reeder, 2021; Fung, Jones & Doshi, 2023; Heying & Weinstein, 2023; Ioannidis, 2022; Fenton & Neil, 2023). One major kind of selection bias takes the form of miscategorisation, whereby study participants who have been vaccinated are miscategorised as unvaccinated up to and until some arbitrarily defined time after vaccination occurred (typically up to 14 or 21 days). This selection bias, which has been seen to take several different types, all of which help exaggerate vaccine efficacy, has recently become known colloquially as the ‘cheap trick’ (Fenton & Neil, 2023).

To identify the different types of miscategorisation bias and evaluate how widespread it is, we conducted a review of Covid-19 vaccine studies to identify those studies that have employed miscategorisation selection bias and we have simulated the effects of this selection bias on measures of vaccine efficacy.

This review reveals that, up to February 2024, 38 research studies on Covid-19 vaccines have employed different types of this bias, with variants including straightforward miscategorisation from one category to another, miscategorising the vaccinated as having unverified vaccination status,

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
uncontrolled reporting of vaccination status and excluding those vaccinated from the study. Many of the studies have applied one or more of these biases within time periods from one week to three.

Our simulation model demonstrated that this selection bias artificially boosts vaccine efficacy in all cases, and with the application of repeated ‘booster’ vaccinations, the efficacy of repeated Covid-19 vaccines could be maintained at artificially high levels in perpetuity. Furthermore, in tandem with this the infection rate would likewise be artificially elevated and would be lower for the unvaccinated cohort compared to the vaccinated cohort, further compounding misleading claims that a Covid-19 vaccine reduces infection rates when it does not.

The paper is structured as follows: In Section 2 we review the work on biases in Covid-19 vaccine studies. In Section 3 we describe the search method by which relevant studies were selected. In Section 4 we classify each of the relevant studies according to novel types of miscategorisation selection biases exhibited. In Section 5 we simulate the vaccine efficacy results that would be observed during peak rollout of both a placebo and negative efficacy vaccine under the various selection biases. Section 6 offers our conclusions.

2. Background

Several studies have investigated bias in Covid-19 vaccine studies, including: (i) outcome reporting bias affecting interpretation of vaccine efficacy where studies report relative risk reduction (RRR) rather than actual risk reduction (ARR) (Brown, 2021); confounding bias in test-negative studies where other acute respiratory infections (ARI) are assumed to occur or be independent to Covid-19 (Doll et al, 2022), where authors promote the use of recently vaccinated individuals as a negative control (Hitchings et al, 2022), due to imperfect sensitivity and/or specificity of the test used to diagnose the disease (Eusebi et al, 2023; Williams et al, 2022); state bias wherein limited uptake, or vaccine hesitancy, is said to occur because the general public prefer domestically produced vaccines over foreign-made (Kobayashi et al, 2021) and alternatively, confirmation bias that causes people to disregard public information and results in the same hesitancy (Malthouse, 2023); self-selection bias where participants who have been vaccinated are more likely to also willingly present for swab collection and testing (Glasziou et al, 2022); and collider stratification bias where rather than the usual approach of reporting the relative risk of the disease, Covid-19, test-negative studies use the recently created alternative approach of reporting the relative risk of infection given a second variable, vaccination (Ortiz-Brizuela et al, 2023). The studies discussed here are approximately evenly divided between those that report biases that have exaggerated factors of vaccine safety and efficacy, and those reporting biases have negatively impacted assessment of these factors and resulting public perception.

We focus explicitly on miscategorisation selection biases, which inevitably exaggerate vaccine efficacy. We identify five types of such bias (defined in detail in Section 4), namely: (i) Miscategorisation (the type most closely associated with the miscategorisation selection bias); (ii) Unverified; (iii) Uncontrolled; (iv) Excluded; and (v) Undefined. Previous work (Ioannidis, 2021; Fung, Jones & Doshi, 2023, Latasher 2024) has largely focused only on miscategorisation, so our review is novel as well as more extensive than previous work. Ioannidis (2021) considers miscategorisation in terms of vaccination self-reporting by participants, the need for investigators to provide definitions for what it means to be vaccinated and whether categorisation as vaccinated occurs immediately after vaccination or after some period, and they discuss the possibility for these definitions to themselves cause miscategorisation of vaccination status. Fung et al (2023) examine this issue in terms of a case-counting window bias, in which investigators do not begin counting cases in the fully vaccinated until the arbitrary period after vaccination had passed. They also found that investigators could apply this period to both the vaccine and placebo arms of their study, or to the vaccine group alone.

3. Method

A search was conducted of PubMed and Scopus seeking literature presenting either a retrospective health records or prospective clinical trial of one or more Covid-19 vaccines with efficacy or safety as an endpoint. The search term used was:

[covid] and [vaccine] and [efficacy] and [safety]
The initial search returned 2,209 results. 476 Duplicates were removed, as well as 1,562 that while discussing or mentioning vaccines for Covid-19 did not present a study of vaccine efficacy or safety and 134 single-page works that were a mix of protocol disclosures and abstracts of results. Of the 37 remaining, one additional paper was excluded because it used different forms of miscategorisation that are out of scope for this study, leaving 34 that provide sufficient detail of the inclusion and exclusion criteria for inclusion in this study. A further 4 papers were identified through citation mining of included papers. Each paper was evaluated for a range of aspects that included the manufacturer and type of vaccine, the control cohort comparator (placebo or unvaccinated), the primary outcomes (prevention of infection, hospitalisation, ICU admission or death), the author’s potential conflicts of interest (declared and undeclared) and whether they included one or more types of miscategorisation selection bias. This work reports on the last of these factors.

4. Types of miscategorisation selection bias

Our review identified the following five types of the miscategorisation selection bias:

(a) Miscategorisation: During the arbitrarily defined period the vaccinated are categorised as unvaccinated, twice vaccinated categorised as single vaccinated, or boosted categorised as twice vaccinated (e.g.: Buchan et al, 2022; Stock et al, 2022).

(b) Unverified: Participants whose vaccination status is unknown or unverified are categorised as unvaccinated (e.g.: Rosenberg et al, 2021; Lyngse et al, 2022b).

(c) Uncontrolled: Participants are allowed to self-administer or self-report their vaccination or infection status, became unblinded or sought vaccination outside the study (e.g.: Angel et al, 2021).

(d) Excluded: Participants who are vaccinated but who become infected or died during the arbitrarily defined period are neither categorised as unvaccinated or vaccinated but are instead simply removed from analysis (e.g.: Tabarsi et al, 2023; Heath et al, 2023);

(e) Undefined: The authors of the study fail to provide definitions for either or both vaccinated and unvaccinated cohorts (e.g.: Bermingham et al, 2023b; Nordstrom et al, 2022).

Table 1 lists the incidence and frequency of use for each type of miscategorisation selection bias in Covid-19 vaccine effectiveness research studies. Use of the arbitrary miscategorisation type was ubiquitous, identified in 100% of the reviewed studies. Further, nearly one-third (31%) also used one or more of the other types of bias.

<table>
<thead>
<tr>
<th>Citation</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e) Defined Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dagan et al (2021)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>14 days</td>
</tr>
<tr>
<td>Haas et al (2021)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>7 days</td>
</tr>
<tr>
<td>Rosenberg et al (2021)</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>14 days</td>
</tr>
<tr>
<td>Thomas et al (2021)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>7 days</td>
</tr>
<tr>
<td>Angel et al (2021)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>7 days</td>
</tr>
<tr>
<td>NSW Health (2021)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>14 days</td>
</tr>
<tr>
<td>Plishvili et al (2021)</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>14 days / 7 days</td>
</tr>
<tr>
<td>Andrews et al (2022)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>28 days</td>
</tr>
<tr>
<td>Buam et al (2022)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>21 days / 14 days</td>
</tr>
<tr>
<td>Buchan et al (2022)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>7 days</td>
</tr>
<tr>
<td>Carazo et al (2022)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>14 days</td>
</tr>
<tr>
<td>Chung et al (2022)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>7 days</td>
</tr>
<tr>
<td>Palinkas et al (2022)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>7 days</td>
</tr>
<tr>
<td>Ferdinands et al (2022)</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>14 days</td>
</tr>
<tr>
<td>Lyngse et al (2022)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>7-15 days</td>
</tr>
</tbody>
</table>
5. Simulation of vaccine effectiveness

We used a deterministic temporal simulation to illustrate the effects of the miscategorisation selection bias on vaccine effectiveness and the reported infection rates for different cohorts, vaccinated and unvaccinated. We simulated a hypothetical vaccination campaign starting at week 1 and completing on week 6 with 85% of the observed population vaccinated by that time.

Here we examine several scenarios showing the effect of a one-week, two-week and three-week selection biases for miscategorisation (a) and exclusion (d) and the effects of repeated vaccination, by boosting, on vaccine efficacy and infection reported rates. Two scenarios present a placebo (zero-efficacy) vaccine, which does not affect infection rates, and compare this with a negative-efficacy vaccine, whereby those vaccinated suffer slightly elevated infection rates compared to the unvaccinated.

Note that observational studies might suffer from many sources of additional confounding biases so this model is a simplification and should not be taken as representative of population level data.

The scenarios simulated cover an eleven-week period with an assumed constant weekly infection rate of 1% in the placebo scenario, and a slightly elevated infection rate, 1.25%, for the vaccinated cohort in the negative-efficacy scenario. This is used in both the miscategorisation, (a), and excluded, (c), simulations. To simulate the effects of boosters we assume a population that is repeatedly vaccinated every twelve weeks, with those who are vaccinated miscategorised (a) within one week of each vaccination.

The results of the five scenarios are presented in Figure 1.

<table>
<thead>
<tr>
<th>Study</th>
<th>Miscategorisation</th>
<th>Exclusion</th>
<th>Time Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lyngse et al (2022b)</td>
<td>X</td>
<td>X</td>
<td>7-15 days</td>
</tr>
<tr>
<td>Nordstrom et al (2022)</td>
<td>X</td>
<td>X</td>
<td>14 days</td>
</tr>
<tr>
<td>Petras et al (2022)</td>
<td>X</td>
<td></td>
<td>14 days</td>
</tr>
<tr>
<td>Robles-Fontan et al (2022)</td>
<td>X</td>
<td></td>
<td>14 days</td>
</tr>
<tr>
<td>Arbel et al (2022)</td>
<td>X</td>
<td></td>
<td>7 days</td>
</tr>
<tr>
<td>Paternina et al (2022)</td>
<td>X</td>
<td></td>
<td>14 days</td>
</tr>
<tr>
<td>Stock et al (2022)</td>
<td>X</td>
<td></td>
<td>21 days / 14 days</td>
</tr>
<tr>
<td>Bermingham et al (2023)</td>
<td>X</td>
<td></td>
<td>21 days</td>
</tr>
<tr>
<td>Yau et al (2023)</td>
<td>X</td>
<td></td>
<td>Until fully vaccinated</td>
</tr>
<tr>
<td>Mitchell et al (2023)</td>
<td>X</td>
<td></td>
<td>14 days</td>
</tr>
<tr>
<td>Tan et al (2023)</td>
<td>X</td>
<td></td>
<td>7 days</td>
</tr>
<tr>
<td>Al Kaabi et al (2023)</td>
<td>X</td>
<td></td>
<td>14 days</td>
</tr>
<tr>
<td>Tabarsi et al (2023)</td>
<td>X</td>
<td>X</td>
<td>14 days</td>
</tr>
<tr>
<td>Heath et al (2023)</td>
<td>X</td>
<td>X</td>
<td>7 days</td>
</tr>
<tr>
<td>Nadeem et al (2023)</td>
<td>X</td>
<td></td>
<td>14 days</td>
</tr>
<tr>
<td>Anez et al (2023)</td>
<td>X</td>
<td></td>
<td>7 days</td>
</tr>
<tr>
<td>Munoz et al (2023)</td>
<td>X</td>
<td></td>
<td>7 days</td>
</tr>
<tr>
<td>Bermingham et al (2023b)</td>
<td>X</td>
<td>X</td>
<td>21 days</td>
</tr>
<tr>
<td>Liu et al (2023)</td>
<td>X</td>
<td></td>
<td>7 days</td>
</tr>
<tr>
<td>Kitano et al (2023)</td>
<td>X</td>
<td></td>
<td>7 days / 14 days</td>
</tr>
<tr>
<td>Polack et al (2020)</td>
<td>X</td>
<td>X</td>
<td>7 days</td>
</tr>
<tr>
<td>Khairullin et al (2022)</td>
<td>X</td>
<td></td>
<td>14 days</td>
</tr>
</tbody>
</table>

| Total | 38 | 4   | 4   | 4   | 2   |
In practice, most studies do not report vaccine efficacy in the initial week(s) (when no cases are categorised as vaccinated) as this would show up as 100% efficacy. However, note that in all scenarios in the first weeks where efficacy would be reported the starting point for efficacy is over 90%.

In scenario A, miscategorisation, (a), with a placebo, high vaccine effectiveness falls towards zero after one, two or three-week periods, accompanied by an increase in the reported infection rate for the unvaccinated cohort from the start of the vaccination campaign. After seven weeks the reported infection rates for the vaccinated and unvaccinated cohorts converge on the true infection rate. In scenario B, miscategorisation, (a), with a negative effectiveness vaccine, the reported vaccine effectiveness is negative from week six onwards, and again the reported infection rate for the unvaccinated is overestimated from the start of the vaccination campaign. However, by the end of the campaign the reported infection rates for the vaccinated would be greater than that for the unvaccinated.

Scenarios C and D are simply the same as scenarios A and B, except for the fact that they are for the excluded type, (c), of selection bias. Note that here the reported infection rate for the unvaccinated remains unbiased whilst that for the vaccinated rises to match the true rate for the placebo and negative efficacy scenarios.

In Scenario E, boosting with miscategorisation, (a), we can see that repeated application of the vaccine at twelve-week intervals restores vaccine efficacy to high levels after each booster and, assuming a constant infection rate, elevates the reported infection rate in the unvaccinated cohort between each booster campaign, giving rise to bias and gross overestimation.

Our simulation model has demonstrated that the effects of this selection bias are to artificially boost vaccine efficacy in all cases, and with the application of repeated ‘booster’ vaccinations, the efficacy of repeated Covid-19 vaccines could be maintained at these artificial levels in perpetuity should boosting be continued indefinitely. Furthermore, in tandem with this the infection rate is likewise artificially elevated for the unvaccinated cohort compared to the vaccinated cohort, further compounding false claims that a Covid-19 vaccine reduces infection rates. Note that other metrics of vaccine effectiveness,
such as mortality or morbidity improvements, are capable of being mis-reported in a similar way because of the same bias.

6. Conclusions

Our reviews reveal that a serious form of selection bias, miscategorisation, is pervasive throughout the many research studies that aim to measure Covid-19 vaccine efficacy. The effect of this bias is to artificially inflate vaccine efficacy and present the misleading impression that these vaccines are effective and that the non-vaccinated suffer from higher Covid-19 infection rates compared to the vaccinated.

We presented a simulation model to demonstrate the effects of this selection bias and show it artificially boosts vaccine efficacy in all cases, and with the application of repeated ‘booster’ vaccinations, the efficacy of repeated Covid-19 vaccines could be maintained at artificial levels in perpetuity should boosting be continued indefinitely. This effect occurs with a both a zero-efficacy (placebo) vaccine and a negative-efficacy vaccine that increases, rather than reduces, infection rates in those vaccinated.

This miscategorisation is guaranteed to lead to initially very high efficacy claims (usually above 90%) during peak vaccine rollout even if the vaccine were a placebo or worse. Efficacy then falls toward zero a few weeks later. This pattern of high initial efficacy, tapering off after 3 months is also consistently observed in real-world studies, and is often used as justification for additional, booster vaccinations to maintain efficacy. The corresponding Covid-19 infection rate is also likewise artificially elevated in the unvaccinated cohort compared to the vaccinated cohort. These issues apply to other measures of vaccination effectiveness related to mortality and morbidity.

Thus, we conclude that any claims of Covid-19 vaccine efficacy based on these studies are likely to be a statistical illusion.

References


Appendix: Research studies containing miscategorisation as a selection bias


Reeder, M. (2021) Use of a null assumption to re-analyze data collected through a rolling cohort subject to selection bias due to informative censoring. Unpublished manuscript. https://zenodo.org/records/5243901


