PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	COVID-19 vaccination effectiveness rates by week and sources of bias: a retrospective cohort study
AUTHORS	Ostropolets, Anna; Hripcsak, George

VERSION 1 – REVIEW

REVIEWER	Henry, David Bond University, CREBP
REVIEW RETURNED	11-Mar-2022

GENERAL COMMENTS	The authors have used data from Columbia University Medical Center data linked to NY State and City Immunization Registries to study historical cohorts of vaccinees who received one or more doses of the three Covid-19 vaccines used commonly in the USA. I think the authors have used appropriate methods to follow their subjects and to adjust for potential confounders. They have used medical chart review to study the behaviors of vaccinated and unvaccinated individuals, have studied the effects of varying endpoint definitions on estimates of vaccine effectiveness and have illustrated the effects of defining different inception dates for follow-up of unvaccinated controls. The main study is unremarkable. This is a statement not a criticism. As of early March, the Johns Hopkins Covid vaccine database contains details of 216 non-randomized studies of effectiveness of Covid-19 vaccines, of which 63 were conducted in the USA (https://view-hub.org/covid-19/effectiveness-studies). These studies vary in their methodological rigor but those that I am familiar with include sophisticated target trial emulation and test negative case-control studies studying infections, transmission, hospitalization, and death. Investigations have addressed temporal effects and sources of bias and the UK Health Security Agency has gone on to examine duration of vaccine protection, effects of third doses and impacts on BA1 and BA2 Omicron sub-variants. 1) My point is that the many existing observational studies have provided a great deal of information about effectiveness of Covid-19 vaccines. The authors need to be clearer about what their study adds to the current body of knowledge about the impact of vaccines. Presently, their conclusions are framed as a motherhood statement about the need to account for biases in observational research 2) The study is limited in not providing contemporary data on how VE declines over time, particularly after 3 doses and against the Omicron sub-variants. The authors acknowledge this but it diminishes the study's currency

	3) The authors do raise important questions about the healthcare-seeking behaviors of vaccinated and unvaccinated subjects. This is relevant to the interpretation of the immunologically implausible observation of very early vaccine protection (within 1 week of vaccination). This effect was not seen in randomised trials but has been noted in other observational studies. I think the authors have highlighted some interesting findings through chart review, but I don't think this discussion is adequately developed in the paper and the potential confounding effects of these differential behaviors are not fully explored. 4) The authors haven't explored the broader literature on post-vaccination health behaviors and how they impact on VE research. It is an important and genuinely complex discussion. And it is likely to have changed over the course of the epidemic as those who now remain unvaccinated represent a resistant group, some with extreme views, who are different from unvaccinated individuals early in the roll-out phases. There are other complexities. In an early Israeli VE study the investigators had to fully adjust their model (including a variable for previous [pre-Covid] vaccine seeking behavior) to counter implausibly early post vaccination protection, which may have been due to a 'healthy vaccinee' effect. From memory, early observational studies in the UK found variable effects (some found very early benefit, others didn't). This may have been influenced by the targeting of the highest risk groups for early vaccination with specific vaccines. In terms of health behaviors, Australian researchers found that unvaccinated individuals were less likely to say they would present for Covid-19 testing than vaccinated individuals if they developed symptoms. I am laboring this point with examples as I think it is a clear focus of this paper and it needs a deeper discussion, which would be a welcome feature of this paper. 5) The authors have used a common data model and standardized analytic approaches to derive measures of vaccine effectiveness. I am a little unclear about how the cohorts were assembled. The vaccination data included individuals across NY state. But the cohorts had to have 365 days prior observation in their hospital system. Does this mean that they were enrolled in CUIMC during that period and what assurance is there that they were continuously enrolled after inception? Could they have presented for care at other hospitals in NY city or state? 6) I have 4 comments about the presentation of methods and results: a. The details in the Methods in the main body of the paper are sparse; For instance, I cannot find mention of censoring rules in the main body or appendices. b. I was unable to read several of the Tables and Figures properly. For instance, Table 1 showing covariate balance before and after propensity score matching may have been corrupted during pdf conversion and the narrow column widths hinders legibility. c. The figures are too small and too busy to be easily read and understand. As an example, what I assume is Figure 3 (it's not labelled as such) the top panels display VE against different endpoints during different follow up periods and the bottom panel displays results of the chart review. I think the latter should be in a separate figure and pie charts are sub-optimal. I think a bar chart

	would be preferable. d. The presentation of KP curves could be improved. The figures are crowded and the text is small and hard to read. The vertical axis is labelled 'probability of survival', which may be misunderstood in a literal sense. Probability of remaining free of the endpoint might be more accurate. In the upper middle KP display what I assume is the 95% CI disappears before the end of the follow up period. There is insufficient detail on numbers at risk and no information on numbers censored during the follow up period. 7) Although these are no deal breakers the language used in different parts of the manuscript is a little odd a. The authors use the words 'absolute effectiveness'. I am unclear why. Generally, in comparative effectiveness research absolute difference is calculated as the difference between two rates. But the vaccine effectiveness estimate they use (1-HR) is a relative measure of effectiveness. So their use of the term 'absolute' is confusing b. They also use the term 'granular weekly effectiveness estimates'. I don't know what this means. 'Weekly' implies a rate – for instance a weekly decline in effectiveness. But what they are describing is the estimated VE during specific time intervals after the vaccination dates. They have also used the term 'week by week' which is also unclear. c. The authors refer to 'large-scale propensity score matching' but I don't know what this means. I don't think the authors have used a high dimensional approach to generating propensity scores – I couldn't find mention in the manuscript. So clarification is needed. d. The authors have used many negative 'control' outcomes. I am not clear about why they chose so many. There doesn't seem to have been a clinical consideration around their selection – they appear to be an alphabetical list of SNOMED codes. I also question the logic of their analysis. Their principal interest is in the effects of exposure – receipt of Covid-19 vaccination. To me it would make more sense to use negative control exposures – e.g., medications that are not thought to influence the chance of getting Covid-19 or its complications. e. I don't understand the difference between initial outcome: "b) COVID-19 hospitalization defined as an inpatient visit associated with a COVID-19 positive test or diagnosis within 30 days prior or during the visit" and additional outcome "b) COVID-19 hospitalization associated with a positive COVID-19 test." f. The authors use the terms 'target' and 'comparator' groups rather than 'exposed' and 'unexposed' or 'vaccinated' and 'unvaccinated'. I think their language may come from the analytics package but may be unfamiliar to readers g. I don't understand most of the section labelled 'Sensitivity Analyses'. It is placed before the statistical analysis section, which is unusual. But my main concern here is the authors' use of the term 'sensitivity analysis'. Generally (but not exclusively) it refers to analyses where the inputs are varied across credible ranges to determine how sensitive the analytical coefficient is to these varying assumptions. I admit the term is often used loosely, but the authors

	here appear to be describing separate analyses that address different questions. For example, they state “we assessed overall absolute vaccine effectiveness in patients with at least one dose of a COVID-19 vaccine and in fully vaccinated patients.” I may have misunderstood, but that’s not a sensitivity analysis. These are analyses of partial and full vaccination. We expect these to produce different results, underpinned by a strong biological rationale. They go on “our second sensitivity analysis included all eligible patients regardless of their previous COVID-19 status. This is asking a separate question about whether vaccine effectiveness varies in a population that includes subjects who have already had Covid-19. Of course, the enhanced immunity after vaccination of previously infected subjects is a very salient question, but it’s not a sensitivity analysis.

REVIEWER	Hyams, Catherine University of Bristol, Academic Respiratory Unit CH is the PI for the AvonCAP study, which is an investigator led surveillance study sponsored by the University of Bristol and funded by Pfizer.
REVIEW RETURNED	22-Mar-2022

GENERAL COMMENTS	MAJOR: 1. I am unsure why the authors highlight American studies in the second paragraph of the introduction, downplay non-US studies yet have submitted this publication to an English publishing house? - I accept that this as an analysis from the US, the points raised in the final sentence of this paragraph are valid. Nonetheless, I feel that there is a distinct flavour of 'American studies are superior to everyone else's' in the opening paragraph which needs to be revised before publication 2. I do not believe that this is a 'test-negative' design case control study, as cases are defined as either test positive or clinical diagnosis in the absence of a positive test a) please can the authors confirm this b) if this is the case, then the methodology needs to be updated to specify that the clinical diagnosis of COVID could occur despite a negative test c) The authors need to explicitly state how many participants are in each of the above groups (ie test positive, clinical diagnosis only, neither test positive nor clinical diagnosis). - This should also be laid out by each vaccine brand and control group please - Also, are these only PCR tests, or were lateral flow tests considered acceptable for diagnosis? If so, this should be made explicitly clear 3) there is no explanation of what happened to patients with discordant COVID test results (ie first test negative, second test positive for same hospital episode, etc). Please provide some clarity here 4) following on from point 2, I would expect to see the analysis repeated (ie a sensitivity analysis), when only PCR positive patients were used as cases and the original control group were used as controls (ie the clinical COVID diagnosis group removed). As is currently presented, I am unsure what bias testing/clinical diagnosis may have on the results of this study

	5) Presumably the comorbidities listed in table 1 allow for the Charlson Comorbidity Index to be calculated. This would be helpful to include please 6) Is there any indication of socioeconomic status of participants to be included? - in a fee-paying or insurance based system, one could reasonably assume that those less well off would be less likely to attend medical appointments or seek healthcare, and I do not see this discussed in the limitations at all 7) I confess to being confused as to whether this paper is attempting a head to head comparison or not? - if it is, then it must be explicit and justify their methodology and power - it it is not, then it must also explicitly state that it isnt and explain why the estimates are not comparable (and rework any figures that have the vaccines combined/alongside each other and allow for direct comparison) 8) I found the Table 1 quite hard to look at and review, as all of the numbers appear over multiple lines due to column sizing.... - also, please can there be some footnotes to explain what met the definition for the different conditions listed in the medical history (eg hypertensive disease - was this any or only that associated with cardiovascular disease? what is renal impairment: any or specific stages of CKD? What is heart disease exactly?) MINOR - I dislike the use of the word 'fast' in line 14. It implies that perhaps not enough time was taken to properly assess vaccine effect or safety. Please can I ask for it to be changed to 'rapid' or some other such word - the limitation section is very limited in outlining the study limitations and needs to be improved - is this study generalisable? what are the biases? how was missing data addressed? what effect does a fee-based/insurance-based system have? where results obtainable by variant and for all vaccine brands, if not why not?

REVIEWER	Olson, Samantha CDC
REVIEW RETURNED	02-Apr-2022

GENERAL COMMENTS	Thank you for allowing me to review this study titled, “COVID-19 vaccination effectiveness rates by week and sources of bias.” This manuscript described vaccine effectiveness (VE) against COVID-19 infection and hospitalization using a large retrospective cohort and assessed bias in observational VE studies. This study adds to the body of literature on COVID-19 VE but is missing some key information to allow sufficient interpretation of these data in the context of other COVID vaccine effectiveness studies. If authors could modify the manuscript, this would significantly improve the contribution. Major comments: • Given observational studies have observed different vaccine effectiveness estimates against different SARS-CoV-2 variants, different age groups, among patients with and without

	immunocompromising conditions, and among other groups, authors need to provide context to the time period of this study, age range of those included, and who these patients are (e.g. % of underlying health conditions, if available, and does the sample include all patients captured by the electronic health system covering both inpatient and outpatient facilities) at the beginning of the manuscript to interpret these effectiveness estimates in context with other published estimates. • Please, provide more detail around the selection of an index date for the unvaccinated patient (methods, page 6, lines 37-40)? An example of date selection might make this more clear. Minor comments: • Are antigen test results also included in the dataset, and if so, how are those results interpreted? • It was likely difficult to collect vaccination history on non-New York residents (methods, page 6, lines 40-42). However, if the electronic medical records or linked registries provide sufficient data on these patient's vaccination histories, authors could further explain this. If out-of-state patients' histories are not well documented, authors may consider removing these patients, quantifying them, or conducting a sensitivity analysis. Otherwise, this may introduce some misclassification of vaccination status. • For the sensitivity analysis around prior infection, was date of prior infection available? It is possible more recent infections would differ than infections that were >90 days from the reference date. It may be helpful to provide some context around dates of prior infection and consider only examining those more recently infected in the sensitivity analysis. • Authors included patients 10-19 years-of-age according to Table 1, but if the study ended in May 2021 (according to Figure 1), vaccination was not available to all in this age group. Suggest removing children not age eligible for vaccination during the time of the study. • Could authors mention IRB approval in methods? It was mentioned that no patients were involved (methods, page 8, line 38), but it would be helpful to be clear that IRB is in-place. • The authors highlight how misclassification of covid-19 cases can occur in VE studies with incidental positives that are seeking care for reasons other than COVID-19. Providing examples of observational studies that use a case definition in addition to a positive test may help support this argument. • Can you provide case and control vaccinated and unvaccinated counts for figure 3 to demonstrate the sufficient sample size for weekly effectiveness calculations.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Prof. David Henry, Bond University

1) My point is that the many existing observational studies have provided a great deal of information about effectiveness of Covid-19 vaccines. The authors need to be clearer about what their study adds to the current body of knowledge about the impact of vaccines. Presently, their conclusions are framed as a motherhood statement about the need to account for biases in observational research
Response: Our main goal was to better understand the biases associated with vaccine effectiveness studies in the setting of COVID-19 or future similar diseases, using the previous COVID-19 vaccine effectiveness literature as a comparison against which we can verify our results (e.g., similar effectiveness after three weeks). In this way, we can show that observational data can be reliably used to estimate long-term COVID-19 effectiveness despite the bias present.

We have made our recommendations more specific, such as reporting early effectiveness rates rather than withholding them, as many recent studies have done. We note that first-week biases may not necessarily invalidate a study's later results, based on both the similarity of our estimates after three weeks to those in the literature and our manual review, which revealed several biases associated specifically with the peri-vaccination period. Nevertheless, we believe that they should be reported in all studies. We expanded our discussion to compare the findings on health-seeking behavior and the previous studies and highlighted the biases observed are likely to vary across different institutions, healthcare settings, countries, and time. Therefore, one should not assume the direction of bias and need to report the results for the first week(s). Knowing what all studies are finding in the first week and separately the second week will help to uncover what is happening in different study designs and more confidently answer what are the implications of the first week results.
We also advise that if a cohort design is employed, the comparison group should be an arbitrary date, based both on our previous measurement study cited in the paper and our current results using that comparator.

2) The study is limited in not providing contemporary data on how VE declines over time, particularly after 3 doses and against the Omicron sub-variants. The authors acknowledge this but it diminishes the study's currency
Response: As you noted before, the aim of this paper was to understand biases associated with vaccination, which are most pertinent to acute changes and not to study long-term vaccine effectiveness. Previous studies on one, two or three doses of COVID-19 vaccines sometimes did not report their findings for the first week(s). We argue that the estimates should be presented for those week(s) and that it provides useful insights on magnitude and direction of bias in different institutions and countries. We modified our background to reflect it (page 4) and noted this in the discussion.

3) The authors do raise important questions about the healthcare-seeking behaviors of vaccinated and unvaccinated subjects. This is relevant to the interpretation of the immunologically implausible observation of very early vaccine protection (within 1 week of vaccination). This effect was not seen in randomised trials but has been noted in other observational studies. I think the authors have highlighted some interesting findings through chart review, but I don't think this discussion is adequately developed in the paper and the potential confounding effects of these differential behaviors are not fully explored.
Response: Many thanks for highlighting this point! We fully agree and expanded discussion so that it highlights comparison with other studies and adds more on the biases we observed.
Specifically, we highlighted that the bias is present even if the analysis adjusts for healthcare utilization and prior vaccination behavior (first paragraph), compared findings on likelihood of testing (second paragraph), discussed differential symptom severity (fourth paragraph) and outcome misclassification (paragraph 6).
We appreciate this comment as it indeed helped to better shape the findings and improve the manuscript.

4) The authors haven't explored the broader literature on post-vaccination health behaviors and how they impact on VE research. It is an important and genuinely complex discussion. And it is likely to have changed over the course of the epidemic as those who now remain unvaccinated represent a resistant group, some with extreme views, who are different from unvaccinated individuals early in the roll-out phases. There are other complexities. In an early Israeli VE study the investigators had to fully adjust their model (including a variable for previous [pre-Covid] vaccine seeking behavior) to counter implausibly early post vaccination protection, which may have been due to a 'healthy vaccinee' effect. From memory, early observational studies in the UK found variable effects (some found very early benefit, others didn't). This may have been influenced by the targeting of the highest risk groups for early vaccination with specific vaccines. In terms of health behaviors, Australian researchers found that unvaccinated individuals were less likely to say they would present for Covid-19 testing than vaccinated individuals if they developed symptoms. I am laboring this point with examples as I think it is a clear focus of this paper and it needs a deeper discussion, which would be a welcome feature of this paper.
Response: Thank you for pointing out these papers! We added a broader comparison with the existing literature. Specifically, similarly to Israeli and other studies, we adjusted for previous vaccine behavior, healthcare utilization, demographics etc., and yet observed the differences in vaccinated and unvaccinated groups, which were only revealed in chart review. We added it to the discussion on

page 10. We also referenced other studies on the likelihood of being tested after vaccination and differential symptom severity (page 10).

5) The authors have used a common data model and standardized analytic approaches to derive measures of vaccine effectiveness. I am a little unclear about how the cohorts were assembled. The vaccination data included individuals across NY state. But the cohorts had to have 365 days prior observation in their hospital system. Does this mean that they were enrolled in CUIMC during that period and what assurance is there that they were continuously enrolled after inception? Could they have presented for care at other hospitals in NY city or state?
Response: A year of prior observation requirement was met by having a continuous observation period in the CUIMC EHR. Since it is an electronic health record data source, there is no explicit enrollment period (as opposed to administrative claims data), so observation periods are constructed based on the first and last healthcare encounter. This means that the patients could seek care at other healthcare institutions. We described this limitation: "Due to observational nature of the study, the data sources may not have complete capture of patient conditions as the patients could seek care outside of the hospital system. While our outcome phenotype algorithms may be subject to measurement error, we provided additional analyses with alternative outcome definitions" (page 12).

6) I have 4 comments about the presentation of methods and results:

a. The details in the Methods in the main body of the paper are sparse; For instance, I cannot find mention of censoring rules in the main body or appendices.
Response: We describe censoring in the Main analysis section. We reformulated the initial statement to make it clearer: "We calculated vaccine effectiveness during six consecutive 7-day intervals after the first dose. Within each interval, patients were followed-up until an outcome, end of the period or death, whichever came earlier" (page 5, line 220-221).

b. I was unable to read several of the Tables and Figures properly. For instance, Table 1 showing covariate balance before and after propensity score matching may have been corrupted during pdf conversion and the narrow column widths hinders legibility.
Response: We reformatted Table 1 for improved readability so that the vaccines are listed one after another and not side-by-side.

c. The figures are too small and too busy to be easily read and understand. As an example, what I assume is Figure 3 (it's not labelled as such) the top panels display VE against different endpoints during different follow up periods and the bottom panel displays results of the chart review. I think the latter should be in a separate figure and pie charts are sub-optimal. I think a bar chart would be preferable.
Response: We split Figure 3 into two figures: Figure 3 showing effectiveness over the first six 7-days intervals and Figure 4 showing the chart review results. We also changed pie chart to bar chart.

d. The presentation of KP curves could be improved. The figures are crowded and the text is small and hard to read. The vertical axis is labelled 'probability of survival', which may be misunderstood in a literal sense. Probability of remaining free of the endpoint might be more accurate. In the upper middle KP display what I assume is the 95% CI disappears before the end of the follow up period. There is insufficient detail on numbers at risk and no information on numbers censored during the follow up period.
Response: Thank you for pointing it out! We incorporated your comments into the new version of Kaplan Meier plots (Appendix 6-8).

7) Although these are no deal breakers the language used in different parts of the manuscript is a little odd
a. The authors use the words 'absolute effectiveness'. I am unclear why. Generally, in comparative effectiveness research absolute difference is calculated as the difference between two rates. But the vaccine effectiveness estimate they use (1-HR) is a relative measure of effectiveness. So their use of the term 'absolute' is confusing
Response: Here, we used the term "absolute effectiveness" to distinguish the effectiveness of COVID-19 vaccines when compared to unvaccinated population as opposed to comparison to other COVID-19 vaccines. To provide more clarity, we removed the word "absolute".

b. They also use the term ‘granular weekly effectiveness estimates’. I don’t know what this means. ‘Weekly’ implies a rate – for instance a weekly decline in effectiveness. But what they are describing is the estimated VE during specific time intervals after the vaccination dates. They have also used the term ‘week by week’ which is also unclear.
Response: We studied vaccine effectiveness over six 7-day intervals after the first dose, which we referred to as granular weekly estimates. We reformulated this term throughout the paper to make it clearer.

c. The authors refer to ‘large-scale propensity score matching’ but I don’t know what this means. I don’t think the authors have used a high dimensional approach to generating propensity scores – I couldn’t find mention in the manuscript. So clarification is needed.
Response: We used large-scale propensity score model, which is briefly an L1 regularized regression that incorporates a large number of covariates including demographics, drug administration and prescription information, previous patient conditions and procedures, etc. simultaneously (up to 54,987 covariates in this study) (1,2). As described in the work by Tian et al., L1-regularization offers propensity score performance superior to the high dimensional propensity score marginal screen. We clarified the use of large-scale propensity score modelling in the Methods section and added citation to the work by Tian et. al and Fortin et. al describing the approach used in the study.

d. The authors have used many negative ‘control’ outcomes. I am not clear about why they chose so many. There doesn’t seem to have been a clinical consideration around their selection – they appear to be an alphabetical list of SNOMED codes. I also question the logic of their analysis. Their principal interest is in the effects of exposure – receipt of Covid-19 vaccination. To me it would make more sense to use negative control exposures – e.g., medications that are not thought to influence the chance of getting Covid-19 or its complications.
Response: We use the term “negative controls” to describe conditions or states known to not be associated with the exposure (COVID-19 vaccines) (3). Negative controls were selected using a standardized process outlined by Voss et al. (4), which included a) automated literature screening to ensure that that there is no Medline abstract where the MeSH terms suggest an association between the vaccines and the condition, that there is no mention of the drug-condition pair on a US Product Label in the “Adverse Drug Reactions” or “Postmarketing” section, and there are no US spontaneous reports suggesting that the pair is in an adverse event relationship (5). We then limited the set to those conditions and states that are used in the CUIMC data source and are not too broad and selected the most commonly used codes presented in the supplements.
In our previous work, we found that patients who receive COVID-19 vaccination on average have fewer medical events and medication prescriptions/administrations on the date of vaccination and up to a year prior compared to the patients indexed on healthcare encounter (6,7). As medication exposures and administrations are coupled with a visit, those patients may be sicker than vaccinated patients and may have higher chances of getting COVID-19 or of getting a more severe form.

e. I don’t understand the difference between initial outcome: “b) COVID-19 hospitalization defined as an inpatient visit associated with a COVID-19 positive test or diagnosis within 30 days prior or during the visit” and additional outcome “b) COVID-19 hospitalization associated with a positive COVID-19 test.”
Response: The former outcome includes hospitalization associated with either a positive test or a COVID-19 diagnosis, while the latter only uses a positive test (COVID-19 diagnosis is not included).

f. The authors use the terms ‘target’ and ‘comparator’ groups rather than ‘exposed’ and ‘unexposed’ or ‘vaccinated’ and ‘unvaccinated’. I think their language may come from the analytics package but may be unfamiliar to readers
Response: While it is a common language across the OHDSI studies, we appreciate you pointing out that these terms may not be familiar to a broader audience. We changed the language accordingly.

g. I don’t understand most of the section labelled ‘Sensitivity Analyses’. It is placed before the statistical analysis section, which is unusual. But my main concern here is the authors’ use of the term ‘sensitivity analysis’. Generally (but not exclusively) it refers to analyses where the inputs are varied across credible ranges to determine how sensitive the analytical coefficient is to these varying assumptions. I admit the term is often used loosely, but the authors here appear to be describing separate analyses that address different questions. For example, they state “we assessed overall absolute vaccine effectiveness in patients with at least one dose of a COVID-19 vaccine and in fully

vaccinated patients." I may have misunderstood, but that's not a sensitivity analysis. These are analyses of partial and full vaccination. We expect these to produce different results, underpinned by a strong biological rationale. They go on "our second sensitivity analysis included all eligible patients regardless of their previous COVID-19 status. This is asking a separate question about whether vaccine effectiveness varies in a population that includes subjects who have already had Covid-19. Of course, the enhanced immunity after vaccination of previously infected subjects is a very salixent question, but it's not a sensitivity analysis.
Response: Sensitivity analyses represent a section where we attempted to distinguish the influence of different factors on vaccine effectiveness, such as a) different index dates for unvaccinated population, b) influence of prior infection, c) different outcome definitions. Finally, we looked at the fully vaccinated patients to compare our estimates to the clinical trials' results. We changed the title of the section to "Secondary analyses" and provided further clarification on why we looked at fully vaccinated patients ("Finally, we assessed vaccine effectiveness in patients with at least one dose of a COVID-19 vaccine and in fully vaccinated patients over all available follow-up to compare the estimates to the results of the RCTs", page 6, line 280-284). We also re-organized the section

Reviewer: 2
Dr. Catherine Hyams, University of Bristol
Comments to the Author:
MAJOR:
1. I am unsure why the authors highlight American studies in the second paragraph of the introduction, downplay non-US studies yet have submitted this publication to an English publishing house?
- I accept that this as an analysis from the US, the points raised in the final sentence of this paragraph are valid. Nonetheless, I feel that there is a distinct flavour of 'American studies are superior to everyone else's' in the opening paragraph which needs to be revised before publication
Response: Thank you for pointing it out! We re-wrote the background section to remove the emphasis from the US-approved vaccines and included more studies from Europe and Asia.

2. I do not believe that this is a 'test-negative' design case control study, as cases are defined as either test positive or clinical diagnosis in the absence of a positive test
a) please can the authors confirm this
b) if this is the case, then the methodology needs to be updated to specify that the clinical diagnosis of COVID could occur despite a negative test
c) The authors need to explicitly state how many participants are in each of the above groups (ie test positive, clinical diagnosis only, neither test positive nor clinical diagnosis).
- This should also be laid out by each vaccine brand and control group please
- Also, are these only PCR tests, or were lateral flow tests considered acceptable for diagnosis? If so, this should be made explicitly clear

Response: We used retrospective cohort design as described on page 5, line 197. The outcome included a COVID-19 diagnosis, which could have been recorded with or without positive test or a positive test ("Outcomes of interest included a) COVID-19 infection defined as a positive COVID-19 test (reverse-transcriptase–polymerase-chain-reaction) or a diagnostic code of COVID-19…").
In terms of the test type, our data source only had PCR test available. Its results are reported as positive, negative or inconclusive and for each patient we collected all tests with a positive result. We changed the sentence describing tests to make it explicit: "…defined as a positive COVID-19 test (reverse-transcriptase–polymerase-chain-reaction assay)…", page 5, line 212-213.
The number of participants in each group (exposed and unexposed and for each brand of COVID-19 vaccine) are recorded in the Table 1 (page 20-23) and we also added the number of outcomes (positive test and diagnosis).

3) there is no explanation of what happened to patients with discordant COVID test results (ie first test negative, second test positive for same hospital episode, etc). Please provide some clarity here
Response: As patients may develop COVID-19 during their hospital stay, we included all patients with a positive test regardless of whether a negative or inconclusive test was observed prior to the positive test and the date of the outcome was set to be the date of the positive test.

4) following on from point 2, I would expect to see the analysis repeated (ie a sensitivity analysis), when only PCR positive patients were used as cases and the original control group were used as

controls (ie the clinical COVID diagnosis group removed).
As is currently presented, I am unsure what bias testing/clinical diagnosis may have on the results of this study
Response: As noted above, only PCR test was available. We changed the sentence describing tests to make it explicit: "…defined as a positive COVID-19 test (reverse-transcriptase–polymerase-chain-reaction assay)…", page 5, line 212-213.

5) Presumably the comorbidities listed in table 1 allow for the Charlson Comorbidity Index to be calculated. This would be helpful to include please
Response: We added Charlson comoborbidity score to Table 1 for each vaccine (page 21-23).

6) Is there any indication of socioeconomic status of participants to be included?
- in a fee-paying or insurance based system, one could reasonably assume that those less well off would be less likely to attend medical appointments or seek healthcare, and I do not see this discussed in the limitations at all
Response: While COVID-19 testing has been free and available to everybody in the study cohort, there may be a difference in how exposed and unexposed group sought care. We added a limitation to page 12-13, line 587-605: "We attempted to address potential differences between exposed and unexposed groups by selecting a large number of covariates in our propensity score model such as number of visits, procedure and drug utilization, prior vaccine behavior, race and others. Nevertheless, we did not have data for social interactions, adherence to preventive measures and policies, which could affect likelihood of COVID-19 infection and testing".

7) I confess to being confused as to whether this paper is attempting a head to head comparison or not?
- if it is, then it must be explicit and justify their methodology and power
- it it is not, then it must also explicitly state that it isnt and explain why the estimates are not comparable (and rework any figures that have the vaccines combined/alongside each other and allow for direct comparison)
Response: For our main analysis, we studied weekly effectiveness of mRNA vaccines (Pfizer-BioNTech and Moderna). For secondary analyses, we studied those vaccines and Janssen vaccine separately and did not attempt a head-to-head comparison. We explicitly argue against it in the Discussion section: "An indirect comparison may not be accurate due to the differences in the populations we observed in our study. First, patients vaccinated with Janssen were substantially different from mRNA patients: on average, they were older, had a higher proportion of patients with race recorded as Black and had more comorbidities. Therefore, comparative effectiveness studies of Janssen and mRNA vaccines require robust techniques such as large-scale propensity matching to ensure valid comparison" (page 12, line 565-567).
We changed the figure Kaplan-Meier plots so that the vaccines are not presented side-by-side (Appendix 6-8).

8) I found the Table 1 quite hard to look at and review, as all of the numbers appear over multiple lines due to column sizing....
- also, please can there be some footnotes to explain what met the definition for the different conditions listed in the medical history (eg hypertensive disease - was this any or only that associated with cardiovascular disease? what is renal impairment: any or specific stages of CKD? What is heart disease exactly?)
Response: We added the footnotes to Table 1 as below:
"* Hypertensive disorder includes primary and secondary hypertension
** Renal impairment includes acute and chronic renal failure (prerenal and renal);
*** Heart disease includes cardiac arrythmias, heart valve disorders, coronary arteriosclerosis, heart failure, cardiomyopathies, etc."

MINOR
- I dislike the use of the word 'fast' in line 14. It implies that perhaps not enough time was taken to properly assess vaccine effect or safety. Please can I ask for it to be changed to 'rapid' or some other such word
Response: We changed the word 'fast' to "rapid" (page 4, line 111).

- the limitation section is very limited in outlining the study limitations and needs to be improved - is

this study generalisable? what are the biases? how was missing data addressed? what effect does a fee-based/insurance-based system have? where results obtainable by variant and for all vaccine brands, if not why not?
Response: We added the following sentences expanding on the limitations:
"The results of the study may not be generalizable to other countries or settings with different vaccine administration practices and policies." and "Finally, the study period did not allow us to stratify the results by COVID-19 variants, which limits generalizability of findings to other variants" (page 13, line 607-609).
We acknowledged a potential issue of incomplete data capture: "Due to observational nature of the study, the data sources may not have complete capture of patient conditions, which was mitigated by having free and available COVID-19 testing and COVID-19 vaccination in Columbia University Irving Medical Center/New York-Presbyterian sites as well as by having data capture from New York City and State Immunization Registries". As noted, testing and vaccination was provided for all adults in New York, which minimizes the influence of insurance-based system. We also acknowledged potential measurement error: "While our outcome phenotype algorithms may be subject to measurement error, we provided additional analyses with alternative outcome definitions" (page 12).

Reviewer: 3
Ms. Samantha Olson, CDC
Major comments:
• Given observational studies have observed different vaccine effectiveness estimates against different SARS-CoV-2 variants, different age groups, among patients with and without immunocompromising conditions, and among other groups, authors need to provide context to the time period of this study, age range of those included, and who these patients are (e.g. % of underlying health conditions, if available, and does the sample include all patients captured by the electronic health system covering both inpatient and outpatient facilities) at the beginning of the manuscript to interpret these effectiveness estimates in context with other published estimates.
Response: We added a brief description of the study sample early on in Results ("In total, we identified 179,666 patients with at least one dose of COVID-19 vaccine in January-May 2021: 121,771 patients for Pfizer-BioNTech, 52,728 for Moderna and 5,167 for Janssen (Table 1). The sample included patients from all age groups, with or without co-morbidities captured in inpatient and outpatient settings") with a link to Table 1 that describes the proportion of patients with specific co-morbidities. We also re-organized the first paragraphs of Results to discuss patient characteristics earlier on in the manuscript.

• Please, provide more detail around the selection of an index date for the unvaccinated patient (methods, page 6, lines 37-40)? An example of date selection might make this more clear.
Response: Given that unvaccinated patients do not have a clear exposure date to be set as their index date (the date they enter the group), we set their index date so that it matched the index date of one of the exposed group participants. We re-wrote the sentence in Methods to make it clearer: "For the unexposed group, we selected unvaccinated patients and set their index date to a date (not necessarily with any medical event) that matched the index date of one of the exposed group participants".

Minor comments:
• Are antigen test results also included in the dataset, and if so, how are those results interpreted?
Response: For a positive test outcome, our data source only had PCR test available. We changed the sentence describing tests to make it explicit: "…defined as a positive COVID-19 test (reverse-transcriptase–polymerase-chain-reaction assay)…", page 5, line 212-213.

• It was likely difficult to collect vaccination history on non-New York residents (methods, page 6, lines 40-42). However, if the electronic medical records or linked registries provide sufficient data on these patient's vaccination histories, authors could further explain this. If out-of-state patients' histories are not well documented, authors may consider removing these patients, quantifying them, or conducting a sensitivity analysis. Otherwise, this may introduce some misclassification of vaccination status.
Response: We agree with this assessment and limited our analysis to only those patients (in both exposed and unexposed group) that reside in New York City as described on page 5, line 153-154. We added a line additionally clarifying exclusion of non-NYC residents.

• For the sensitivity analysis around prior infection, was date of prior infection available? It is possible more recent infections would differ than infections that were >90 days from the reference date. It may be helpful to provide some context around dates of prior infection and consider only examining those more recently infected in the sensitivity analysis.
Response: Overall, median and interquartile range of the difference in days between a prior episode and a current COVID-19 episode was 82 (41 – 154) days if an episode was defined as a COVID-19 diagnosis or a positive test. Given that a COVID-19 diagnosis may sometimes be assigned to code sequelae of disease, we also computed the difference between a prior episode and a current episode defined as a positive test, which was 34 (22 – 57) days.

• Authors included patients 10-19 years-of-age according to Table 1, but if the study ended in May 2021 (according to Figure 1), vaccination was not available to all in this age group. Suggest removing children not age eligible for vaccination during the time of the study.
Response: Indeed, the vaccination was not necessarily available for younger patients at the time. As seen in the Table 1, the proportion of patients under 19 is on average less than 2% and this age group is mainly present for Pfizer vaccine, which was received by some of the younger patients. Given small proportion of such patients and the fact that our propensity score model was able to achieve balance on age (i.e., we included similar proportion of patients under 19 in vaccinated and unvaccinated groups) we believe that their inclusion has little if any impact on study estimates.
• Could authors mention IRB approval in methods? It was mentioned that no patients were involved (methods, page 8, line 38), but it would be helpful to be clear that IRB is in-place.
Response: The IRB approval is specified in the Ethical Approval section (page 14, “The protocol for this research was approved by the Columbia University Institutional Review Board (AAAO7805)”).

• The authors highlight how misclassification of covid-19 cases can occur in VE studies with incidental positives that are seeking care for reasons other than COVID-19. Providing examples of observational studies that use a case definition in addition to a positive test may help support this argument.
Response: We added a couple of examples of the papers that used ICD-10 or ICD-10(CM) diagnostic codes for identifying patients with COVID-19 in COVID-19 vaccine effectiveness studies on page 11, line 519-529. We additionally added a sentence that further discusses the findings on outcome misclassification: “Some researchers have previously reported high positive predictive value of ICD-10 diagnostic codes for COVID-19, which points out that index date misclassification should be scrutinized in each institution participating in the analysis to make valid inferences (8,9)”.

• Can you provide case and control vaccinated and unvaccinated counts for figure 3 to demonstrate the sufficient sample size for weekly effectiveness calculations.
Response: We added the counts to the following sentence: “Figure 3, A shows week-by-week estimates for patients vaccinated with at least one dose of Pfizer-BioNTech or Moderna (160,114 patients) compared to unvaccinated patients (115,689).”

References:
1. Tian Y, Schuemie MJ, Suchard MA. Evaluating large-scale propensity score performance through real-world and synthetic data experiments. International Journal of Epidemiology. 2018 Dec 1;47(6):2005–14.
2. Fortin SP, Johnston SS, Schuemie MJ. Applied comparison of large-scale propensity score matching and cardinality matching for causal inference in observational research. BMC Med Res Methodol. 2021 Dec;21(1):174.
3. Schuemie MJ, Ryan PB, Hripcsak G, Madigan D, Suchard MA. Improving reproducibility by using high-throughput observational studies with empirical calibration. Phil Trans R Soc A. 2018 Sep 13;376(2128):20170356.
4. Voss EA, Boyce RD, Ryan PB, van der Lei J, Rijnbeek PR, Schuemie MJ. Accuracy of an automated knowledge base for identifying drug adverse reactions. Journal of Biomedical Informatics. 2017 Feb;66:72–81.
5. Banda JM, Evans L, Vanguri RS, Tatonetti NP, Ryan PB, Shah NH. A curated and standardized adverse drug event resource to accelerate drug safety research. Sci Data. 2016 Dec;3(1):160026.
6. Ostropolets A, Ryan PB, Schuemie MJ, Hripcsak G. Differential anchoring effects of vaccination comparator selection: characterizing a potential bias due to healthcare utilization in COVID-19 versus influenza [Internet]. Epidemiology; 2021 Oct [cited 2021 Nov 7]. Available from: http://medrxiv.org/lookup/doi/10.1101/2021.10.07.21264711
7. Ostropolets A, Li X, Makadia R, Rao G, Rijnbeek PR, Duarte-Salles T, et al. Factors Influencing

Background Incidence Rate Calculation: Systematic Empirical Evaluation Across an International Network of Observational Databases. Front Pharmacol. 2022 Apr 26;13:814198.
8. Bodilsen J, Leth S, Nielsen SL, Holler JG, Benfield T, Omland LH. Positive Predictive Value of ICD-10 Diagnosis Codes for COVID-19. CLEP. 2021 May;Volume 13:367–72.
9. Lynch KE, Viernes B, Gatsby E, DuVall SL, Jones BE, Box TL, et al. Positive Predictive Value of COVID-19 ICD-10 Diagnosis Codes Across Calendar Time and Clinical Setting. CLEP. 2021 Oct;Volume 13:1011–8.

VERSION 2 – REVIEW

REVIEWER	Henry, David Bond University, CREBP
REVIEW RETURNED	21-Jun-2022

GENERAL COMMENTS	I think the authors have cleverly edited the article to emphasise the methodological aspects that are generalisable to other studies attempting to estimate vaccine effectiveness. There are a few minor editing issues to clean up. The words 'sensitivity analyses' remain in the Abstract - 'secondary analyses' are used in the text. Table 1 retains 'target' and 'comparator' as column headings, rather than 'vaccinated' and 'unvaccinated'. There are a couple of sentences that need clarification. In the Abstract, I think the wording referring to the finding of very early vaccine effectiveness needs work. Its a tough concept to summarise in a sentence. My effort is 'Effectiveness in the first week(s) after the vaccination should be reported, as studies are capable of accurately estimating longer term vaccine effectiveness in the presence of implausible early results.' I am sure the authors can improve on this. Under 'Main Design' lines 141/142 they state 'The data were translated to the OMOP Common Data Model version 5 and was previously used in multiple studies (27).' Should this read ' as was used in multiple studies?'

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
I think the authors have cleverly edited the article to emphasise the methodological aspects that are generalisable to other studies attempting to estimate vaccine effectiveness. There are a few minor editing issues to clean up. The words 'sensitivity analyses' remain in the Abstract - 'secondary analyses' are used in the text. Table 1 retains 'target' and 'comparator' as column headings, rather than 'vaccinated' and 'unvaccinated'. There are a couple of sentences that need clarification. In the Abstract, I think the wording referring to the finding of very early vaccine effectiveness needs work. Its a tough concept to summarise in a sentence. My effort is 'Effectiveness in the first week(s) after the vaccination should be reported, as studies are capable of accurately estimating longer term vaccine effectiveness in the presence of implausible early results.' I am sure the authors can improve on this. Under 'Main Design' lines 141/142 they state 'The data were translated to the OMOP Common Data Model version 5 and was previously used in multiple studies (27).' Should this read ' as was used in multiple studies?'

Response: Many thanks for your feedback again; your previous comments helped greatly in shaping a better paper.

We changed sensitivity analyses to secondary throughout the paper and the abstract and changed target and comparator to vaccinated and unvaccinated in Table 1. We modified the sentence on using OMOP Common Data Model accordingly.

Thank you for your suggestion on improving the Abstract, we indeed thought for a while on how to summarize the early effectiveness findings and recommendations. We modified it as follows: “While we found that studies may be capable of accurately estimating long-term effectiveness despite bias in early weeks, the early week results should be reported in every study so that we may gain a better understanding of the biases”.