

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	EFFECT OF VACCINATION ON THE CASE FATALITY RATE FOR COVID-19 INFECTIONS 2020 – 2021: Multivariate modeling of data from the US Department of Veterans Affairs
AUTHORS	Murata, Glen H; Murata, Allison E; Perkins, Douglas J.; Campbell, Heather M.; Mao, Jenny T; Wagner, Brent; McMahon, Ben; Hagedorn, Curt H

VERSION 1 – REVIEW

REVIEWER	Khajanchi , Subhas Presidency University, Mathematics
REVIEW RETURNED	15-May-2022

GENERAL COMMENTS	<p>The COVID-19 pandemic has already spread throughout the world and the people are aware about the diseases and they are using precautions about the pandemic. But, still the covid-19 is spreading very quickly. There are major comments before considering the second round revision.</p> <p>---- The abstract is a little thin and does not quite convey the vibrancy of the findings and the depth of the main conclusions. The authors should please extend this somewhat for a better first impression.</p> <p>---- The manuscript lacks motivation. Author needs to include the motivation of the study.</p> <p>----There is already an abundance of modeling studies on COVID-19, vaccinations, and the months or years to come. However, apart from Ferguson's (now classic) work, Moore and Giordano, very little is said about similar modeling works. This is an issue for three reasons. First, the intended audience for such pieces is made of policy-makers and the general public: they are already facing an abundance of (occasionally conflicting) findings from models. If there is no attempt to contextualize the findings from this piece among others, then we're more likely to be adding noise to a crowded space, instead of providing valuable guidance. Second, several of the modeling assumptions made here may be in line with other pieces (which may provide some strength to the methods) or may be rather unique (which may need more discussion). Finally, as a piece of scientific literature, the contributions should be situated based on what already exists. In sum, the authors should explain how each of their assumptions and modeling choices compares to the literature; how their findings compare to the literature; and hence what is their specific contribution. Related models include, but are not limited to:</p> <p>--https://doi.org/10.1016/j.rinp.2021.104285 --https://doi.org/10.1007/s12190-021-01507-y --https://doi.org/10.1016/j.chaos.2020.110173 --https://doi.org/10.1016/j.chaos.2020.110049 -- https://doi.org/10.1140/epjp/s13360-021-01997-6</p> <p>----Some references contain errors and inconsistent formatting. It is difficult to give credit to research if even elementary aspects of the work are not error free. This should be corrected with care and love to detail.</p> <p>----The manuscript is comprehensive, and I have enjoyed learning about the presented results. I find that the manuscript is written with very poor english and the presentation is not good, and I am in principal in favor of publication, although the following comments should nevertheless be accommodated in one major revision.</p>
-------------------------	---

REVIEWER	Giabbanelli, Philippe Miami University, Computer Science & Software Engineering
-----------------	--

REVIEW RETURNED	21-May-2022
GENERAL COMMENTS	<p>The authors have used a large and detailed dataset (about 340,000 patients from 130 medical centers) to investigate two hypotheses: (H1) vaccination reduces the risk of death of patients with COVID-19; and (H2) the benefits of vaccination continue over time. Through a meticulous statistical analysis, both hypotheses were confirmed. The study is obviously relevant to BMJ Open and supported by three key strengths: the dataset, the number of patient-related aspects used as controls, and the depth of the analysis. Despite these strengths, a few improvements can be made as detailed below.</p> <p>This study was primarily about the Delta variant in the USA, based on data up to September 2021. The title should be amended to reflect some of this time window. The evidence base for COVID-19 moves quickly and, without flagging the specific time window of this study, its findings may be erroneously applied towards another variant such as omicron by readers mostly looking for numbers.</p> <p>Instead of "Large number of COVID patients with complete medical records", the authors may consider "Large number of COVID patients from many medical centers". While medical records are complete, routine laboratory tests often had missing data so I wouldn't emphasize the notion of 'completeness' as a strength here.</p> <p>The introduction needs improvements, in at least three specific ways:</p> <ul style="list-style-type: none"> - the "P(exposure, disease)" and various other P's are unusual and unnecessary, since these probabilities are not actually used in the rest of the document. The manuscript would be more impactful if these P's were removed and the introduction was normally written. If the authors wish to insist on the chain of events that needs to happen, then a visual (classic compartmental model) can easily serve that purpose. - Starting with "Measuring effectiveness in observational", it is no longer the introduction. It becomes the Methods section, as the authors tell us about their procedures. Consequently, the Methods section should start there. - The contributions of the study should be clear in light of other works. <p>COVID-19 has resulted in (perhaps too many) papers, so references are easy to come by. Having only nine references in a paper on COVID-19 vaccination and fatality is very unusual. Most such papers are in the 40-50 references, not as fillers but because a lot has indeed been written and should be cited to contextualize the novelty of this study (for the introduction) or contrast its findings (for the discussion). Some new references are now available (and were not accessible at the time of submission), so the revision can account for them, such as "The impact of COVID-19 vaccines on the Case Fatality Rate" (Int J Infectious Diseases 119), which would be a key comparison point. Other relevant papers could be https://academic.oup.com/ofid/article/9/5/ofac087/6530634 or https://www.degruyter.com/document/doi/10.1515/openhe-2022-0007/html or https://www.sciencedirect.com/science/article/pii/S1567134821004627. This list is certainly not exhaustive, as it is the authors' responsibility to provide this coverage.</p> <p>The paper at times risks to be difficult to read for the audience of BMJ Open. There are very specific techniques as well as medical terms, so it restricts the readership to those having expertise in both domains. I would encourage the authors in making the manuscript more aligned with the intended audience by defining terms. For example, a reader may struggle with a sentence such as "The CSDR was also interrogated for the 2-year (Charl2Yrs) and lifetime Charlson Comorbidity Index (CharlEver) and the 2-year (Elix2Yrs) and lifetime Elixhauser (ElixEver) scores." (This is just one of many such examples.) An alternative to dropping definitions everywhere could be to use a table to list and define the covariates.</p> <p>A figure showing the workflow of the study (data sources, pre-processing such as removing duplicates, analysis, performance measures) would be appreciated.</p>

	<p>For the vaccine cohorts, a histogram (population size per cohort) would be helpful information. This way we can more rigorously appreciate why ≥ 34 weeks were removed and others were not, instead of leaving it to the vagaries of 'too small'.</p> <p>A Table of demographics could replace most (or even all) of the first paragraph of the Results.</p> <p>Results state that 70.5% of participants had complete data for multivariate modeling. And the discussion speaks of a 'conservative approach' that avoids looking at vaccination benefits for those without complete data. So could the methods be clear that patients with missing data were dropped from the analysis?</p> <p>I appreciate the intent, but I have to disagree with the formulation of "Our study is unique in that we performed a systematic review of major domains in the medical record". A systematic review has clear criteria that would be disclosed and can be replicated by others, hence arriving at exactly the same set of variables. Here, we know which aspects are controlled, but we do not necessarily know why these specific aspects were picked and not others. The study is thus very 'comprehensive' in including major domains in the medical record, but perhaps not akin to a 'systematic review'.</p> <p>The statement that "one cannot definitively prove that vaccination improved survival" is not a good idea for public communication. The risk is that this statement becomes isolated and quoted from the article, thus discarding any such study. Rather, the intended spirit of the sentence may be conveyed as, e.g. "there are still many other variables that potentially impact survival."</p> <p>The "Patient and Public Involvement" paragraph essentially boils down to "Not applicable as a secondary data analysis". The rest of the paragraph does not appear to be needed or in the right place (e.g., goal of the study).</p> <p>The analysis script in Stata MP 17 should be provided now, on a permanent repository. It is a normal practice in modeling studies that scripts are provided at the time of review, so they can be reviewed alongside the rest of the material. While sharing patient data is limited by policies and procedures, scripts are not subject to such limitations and should be deposited now, with the URL stated in the manuscript.</p> <p>Minor typos: "possible so included all" -> "possible so we included all" "by the area under its receiver" -> we usually say "the" rather than "its"</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer #1, Dr. Subhas Khajanchi, Department of Mathematics, Presidency University.

- The abstract is a little thin and does not quite convey the vibrancy of the findings and the depth of the main conclusions. The authors should please extend this somewhat for a better first impression.

Response: The abstract was revised according to the Reviewer's suggestion (while remaining within the 300-word limit). A major difference in our analysis of vaccine effectiveness is now highlighted in the abstract conclusion: "Our CFR model controls for the severity of confounding factors and priority of vaccination, rather than using only the presence of comorbidities."

- The manuscript lacks motivation. Author needs to include the motivation of the study.

Response: Our objective was to use the vast, quality data from patients cared for within the Veteran Health Administration to examine the benefits of vaccination (this is stated in the abstract and several other sections of the report). This is also emphasized in the Design subsection the revised manuscript.

- There is already an abundance of modeling studies on COVID-19, vaccinations, and the months or years to come. However, apart from Ferguson's (now classic) work, Moore and Giordano, very little is said about similar modeling works. This is an issue for three reasons. First, the intended audience for such pieces is made of policy-makers and the general public: they are already facing an abundance of (occasionally conflicting) findings

from models. If there is no attempt to contextualize the findings from this piece among others, then we're more likely to be adding noise to a crowded space, instead of providing valuable guidance. Second, several of the modeling assumptions made here may be in line with other pieces (which may provide some strength to the methods) or may be rather unique (which may need more discussion). Finally, as a piece of scientific literature, the contributions should be situated based on what already exists. In sum, the authors should explain how each of their assumptions and modeling choices compares to the literature; how their findings compare to the literature; and hence what is their specific contribution.

Response: The Discussion adds the references suggested by both Reviewers and additional citations relevant to our findings were added throughout the revised manuscript.

- Some references contain errors and inconsistent formatting. It is difficult to give credit to research if even elementary aspects of the work are not error free. This should be corrected with care and love to detail.

Response: The references and formatting were revised according to the BMJ format.

- The manuscript is comprehensive, and I have enjoyed learning about the presented results. I find that the manuscript is written with very poor English and the presentation is not good, and I am in principal in favor of publication, although the following comments should nevertheless be accommodated in one major revision.

Response: The original manuscript scored well on Grammarly Editor, with readability on the level for individuals with some college education. The manuscript has been revised to address this issue and a Grammarly Editor analysis showed a significantly improved overall score.

Reviewer #2, 2 Dr. Philippe Giabbanelli, Miami University.

- This study was primarily about the Delta variant in the USA, based on data up to September 2021. The title should be amended to reflect some of this time window.

Response: The title was amended per this recommendation.

- The evidence base for COVID-19 moves quickly and, without flagging the specific time window of this study, its findings may be erroneously applied towards another variant such as omicron by readers mostly looking for numbers.

Response: The title has been revised accordingly.

- Instead of "Large number of COVID patients with complete medical records", the authors may consider "Large number of COVID patients from many medical centers".

Response: This was done.

- The introduction needs improvements, in at least three specific ways:

- ♣ the "P(exposure, disease)" and various other P's are unusual and unnecessary, since these probabilities are not actually used in the rest of the document. The manuscript would be more impactful if these P's were removed and the introduction was normally written. If the authors wish to insist on the chain of events that needs to happen, then a visual (classic compartmental model) can easily serve that purpose.

- ♣ Starting with "Measuring effectiveness in observational", it is no longer the introduction. It becomes the Methods section, as the authors tell us about their procedures. Consequently, the Methods section should start there.

Response: We agree with the Reviewer's recommendations regarding the acronyms and abbreviations. The manuscript was revised accordingly.

Our paper reports data regarding the case fatality rates for fully vaccinated patients from 2020 to 2021, not an insistence on a chain of events; these data may lend to examination of compartment models, but the primary intent was never to develop such a model. Nonetheless, the Reviewer's suggestion is an excellent one, and we will work with Dr. James Degnan, Ph.D., University of New Mexico Department of Mathematics and Statistics in developing an SAIQR model.

Per the recommendation, the Methods heading now starts with the sentence suggested by the reviewer.

- The contributions of the study should be clear in light of other works. COVID-19 has resulted in (perhaps too many) papers, so references are easy to come by. Having only nine references in a paper on COVID-19 vaccination and fatality is very unusual. Most such papers are in the 40-50 references, not as fillers but because a lot has indeed been written and should be cited to contextualize the novelty of this study (for the introduction) or contrast its findings (for the discussion). Some new references are now available (and were not accessible at the time of submission), so the revision can account for them, such as "The impact of COVID-19 vaccines on the Case Fatality Rate" (Int J Infectious Diseases 119), which would be a key comparison point. This list is certainly not exhaustive, as it is the authors' responsibility to provide this coverage.

Response: These references are excellent suggestions as they highlight how understanding of the severe acute respiratory syndrome coronavirus and public perceptions vacillated since November 2019.

The paper by RK Rai et al, Impacts of social media advertisements on the transmission dynamics of COVID-19 pandemic in India (J Appl Mathematics and Computing 2022) is notable. The Rai et al paper is now cited in the revision.

Samui et al note, "In the absence of any pharmaceutical measures, the public must have to obey the government rules or public health care policies to mitigate the spread of novel coronavirus." This Samui paper is now cited in the revision.

Each of the recommendations have been incorporated into the revised manuscript.

- The paper at times risks to be difficult to read for the audience of BMJ Open. There are very specific techniques as well as medical terms, so it restricts the readership to those having expertise in both domains. I would encourage the authors in making the manuscript more aligned with the intended audience by defining terms. For example, a reader may struggle with a sentence such as "The CSDR was also interrogated for the 2-year (Charl2Yrs) and lifetime Charlson Comorbidity Index (CharlEver) and the 2-year (Elix2Yrs) and lifetime Elixhauser (ElixEver) scores." (This is just one of many such examples.) An alternative to dropping definitions everywhere could be to use a table to list and define the covariates.

Response: We agree with the reviewer. The revision uses clauses that avoid acronyms and abbreviations.

- A figure showing the workflow of the study (data sources, pre-processing such as removing duplicates, analysis, performance measures) would be appreciated.

Response: A supplementary figure depicting the workflow is now included.

- For the vaccine cohorts, a histogram (population size per cohort) would be helpful information. This way we can more rigorously appreciate why ≥ 34 weeks were removed and others were not, instead of leaving it to the vagaries of 'too small'.

Response: Our vaccine effectiveness study was confined to those with the delta variant. Cohorts were defined by time, i.e., in 4-week intervals. As expected, the most recent cohorts are much smaller than the older ones. This can be explained by the fact that there was a marked drop off in the number of veterans seeking vaccination in the summer compared to the previous winter. These cases were not prioritized by VA's triage protocols and probably selected themselves for later vaccination because they considered themselves healthier. Beyond 34 weeks, Effectively, beyond 34 weeks was December 2020 and (other than clinical trials and foreign inoculation – trivial) – no subjects were eligible to be in a cohort prior to at least December 28th, 2020. A supplementary figure is added:

- A Table of demographics could replace most (or even all) of the first paragraph of the Results. Results state that 70.5% of participants had complete data for multivariate modeling. And the discussion speaks of a 'conservative approach' that avoids looking at vaccination benefits for those without complete data. So could the methods be clear that patients with missing data were dropped from the analysis?

Response: A table is now provided. Major strengths of the clinical data in the VA Informatics and Computing Infrastructure are they are comprehensive and longitudinal. If individuals did not have one or more pre-existing conditions, then such individuals were not included in the analysis. (Conversely, having one or more conditions then decreases the likelihood of missing data.) Given the nature of this study, those with missing data were not included in the analysis from the start.

- I appreciate the intent, but I have to disagree with the formulation of "Our study is unique in that we performed a systematic review of major domains in the medical record". A systematic review has clear criteria that would be disclosed and can be replicated by others, hence arriving at exactly the same set of variables. Here, we know which aspects are controlled, but we do not necessarily know why these specific aspects were picked and not others. The study is thus very 'comprehensive' in including major domains in the medical record, but perhaps not akin to a 'systematic review'.

Response: The sentence was removed.

- The statement that "one cannot definitively prove that vaccination improved survival" is not a good idea for public communication. The risk is that this statement becomes isolated and quoted from the article, thus discarding any such study. Rather, the intended spirit of the sentence may be conveyed as, e.g. "there are still many other variables that potentially impact survival."

Response: The sentence was revised to include the suggested clause.

- The "Patient and Public Involvement" paragraph essentially boils down to "Not applicable as a secondary data analysis". The rest of the paragraph does not appear to be needed or in the right place (e.g., goal of the study).

- ***Editor's note regarding reviewer's comment on Patient and Public Involvement - remove the first two sentences, which provide background/rationale that would be better to include in the Introduction. But the rest of the statement is fine as it is. ***

Response: We revised this paragraph in accordance with the Reviewer's and the Editors' suggestions.

- The analysis script in Stata MP 17 should be provided now, on a permanent repository. It is a normal practice in modeling studies that scripts are provided at the time of review, so they can be reviewed alongside the rest of the material. While sharing patient data is limited by policies and procedures, scripts are not subject to such limitations and should be deposited now, with the URL stated in the manuscript.

Response: We have a URL for our affiliated University of New Mexico digital repository (<https://digitalrepository.unm.edu/kinm/>). A request has been initiated to upload the files requested by the reviewer (and this will be publicly accessible within days of this writing).

• Minor typos: "possible so included all" -> "possible so we included all"
 "by the area under its receiver" -> we usually say "the" rather than "its"
 Response: these corrections were made.

Ms. Martin and Reeves, we thank each of you and your Reviewers for the questions and suggestions to improve our report. Every contributor has read and participated in editing this submitted version and this response to the Reviewers. On behalf of the authors,

Brent Wagner, MD, FACP, FASN
 Director, Kidney Institute of New Mexico
 Associate Professor of Medicine
 Division of Nephrology University of New Mexico Health Science Center
 Staff Physician and Associate Chief of Staff for the Research & Development Service
 Medical Section Raymond G. Murphy Memorial Hospital
 New Mexico Veterans Administration Health Care System (501/111c)
 1501 San Pedro Dr. SE
 Albuquerque, NM 87108

VERSION 2 – REVIEW

REVIEWER	Giabbanelli, Philippe Miami University, Computer Science & Software Engineering
REVIEW RETURNED	13-Aug-2022

GENERAL COMMENTS	<p>The authors have reasonably addressed the points that I raised in my review. I note that the supplementary files have now been uploaded (https://digitalrepository.unm.edu/kinm/4/#attach_additional_files), but readers may struggle to find them since the URL is only provided in the response letter. I would encourage the authors in including the URL (now that it is complete) in the main document, in line with best practices for the curation of research data.</p> <p>I appreciate that the authors' revision "uses clauses that avoid acronyms and abbreviations." Unless I am mistaken, I still see several variable names (e.g., PDeathDx, PDeathLabs, AggRiskRx, Charl2Yrs, CharlEver, Elix2Yrs, ElixEver) that may not be clear to the readership. I would thus suggest a brief explanation for the copy editing stage of the manuscript.</p>
-------------------------	--

VERSION 2 – AUTHOR RESPONSE

Reviewer 1: There is already an abundance of modeling studies on COVID-19, vaccinations, and the months or years to come...In sum, the authors should explain how each of their assumptions and modeling choices compares to the literature; how their findings compare to the literature; and hence what is their specific contribution. Response: The Discussion adds the references suggested by both Reviewers and additional citations relevant to our findings were added throughout the revised manuscript.

Editor: We could not find where in the Discussion this comment was addressed, please ensure this comment has been fully addressed and clarify where in the Discussion revisions have been made.

Response. The Discussion section of the Revision contains a fair amount of material regarding assumptions, modeling choices, and the contribution of this submission. All references suggested by both reviewers were added and more. The additions to the revision were,

o Page 16, first paragraph of the discussion: "In this study, we stress the importance of a robust system for classifying vaccine effects, including the severity of confounding factors and the priority of vaccination..."

o Page 18: “The duration of long-lasting T cell responses is in keeping with and may account for our findings demonstrating that the benefit of vaccination on CFR is sustained for at least 30-34 weeks. It is conceivable that while specific components of the immune responses may wane over time, thereby increasing the risk of breakthrough infection, the population of memory T cells recognizing the virus in vaccinated individuals remains relatively stable for at least eight months. As such, when vaccinated individuals are exposed to the SARS-CoV-2 virus, specific memory T cells quickly reactivate, expand in numbers, and rapidly elicit host defense mechanisms capable of mitigating the risk of death from COVID-19 infection.”

To fulfill the response for the Editor, on page 19 we added,

• “To our knowledge, this is the only publication that reports characteristics potentially relevant to mortality in immunized United States Veteran patients during the delta portion of the 2020 pandemic. The multivariate models described herein is unique in employing United States Veteran patient data in a multivariate model (age, a combination of vital signs/laboratory values, Charlson comorbidity indices, pharmaceuticals, and Elixhauser comorbidity scores).”

Reviewer: 2

Dr. Philippe Giabbanelli, Miami University

Comments to the Author:

The authors have reasonably addressed the points that I raised in my review. I note that the supplementary files have now been uploaded (https://digitalrepository.unm.edu/kinm/4/#attach_additional_files), but readers may struggle to find them since the URL is only provided in the response letter. I would encourage the authors in including the URL (now that it is complete) in the main document, in line with best practices for the curation of research data. I appreciate that the authors' revision "uses clauses that avoid acronyms and abbreviations." Unless I am mistaken, I still see several variable names (e.g., PDeathDx, PDeathLabs, AggRiskRx, Charl2Yrs, CharLEver, Elix2Yrs, ElixEver) that may not be clear to the readership. I would thus suggest a brief explanation for the copy editing stage of the manuscript.

Response. The remnant abbreviations were clarified throughout the manuscript as suggested.

- Page 9/10: “The predictor of interest was prior vaccination for COVID-19. Covariates included age, gender, race, ethnicity, veteran status, current smoking, use of supplemental oxygen, probability of death/diagnosis (PDeathDx), laboratory-derived death probability (PDeathLabs), drug class protective effect (AggRiskRx), Charlson comorbidities (2-year/lifetime), Elixhauser comorbidity score (2-year/lifetime), and infection with the delta versus earlier variants.”
- Page 11: “... probability of death/diagnosis (PDeathDx), laboratory-derived death probability (PDeathLabs), and Charlson comorbidities (2-year/lifetime).”
- Page 13: “Probability of death/diagnosis (PDeathDx), laboratory-derived death probability (PDeathLabs), drug class protective effect (AggRiskRx),...”

Every author has seen the revised version of the manuscript and this Response. We are looking forward to seeing the final version in your publication.