Response to Reviewer and Editor Comments:

Comments from Editors

Comment: We've recently seen a paper that recommended against using hospitalised cases of Omicron as the outcome (https://www.sciencedirect.com/science/article/pii/S0264410X22005230). Could you address some of the matters raised in this paper in your discussion?

Response: We agree that including “incidental COVID” cases could erroneously lower estimates of VE, and that this issue is of particular concern in the Omicron era when infections were more common and generally milder. To address this, we considered as ineligible those patients who had no clinical or diagnostic evidence of COVID-like-illness (CLI) other than a positive SARS-CoV-2 test result. This feature of the study design was added to the Methods and this limitation was added to the Discussion:

“Sixth, although we interpret our analyses of the hospital sample as pertaining to “severe” COVID, some hospitalized patients may have tested positive while being hospitalized for other reasons, especially during the Omicron era. To address this, we considered as ineligible patients with no diagnoses suggesting COVID infection other than a positive SARS-CoV-2 test.”

Comment: We would like to take you up on your offer to add additional information. It would be particularly useful if you are able to add information about the 4th dose.

Response: We have updated the manuscript to include encounters through July 12, 2022. In addition, we have added VE estimates for 4th dose recipients.

Comments from Reviewers

Reviewer: 1

The authors applied the test-negative design to linked laboratory, vaccination, and health administrative databases to evaluate waning of 2-dose and 3-dose vaccine effectiveness of mRNA COVID-19 vaccines against hospitalizations and emergency department/urgent care visits for COVID illness in various subgroups (by age group, by immunocompromise, by vaccine product) during periods of Omicron, Delta, and pre-Delta variant predominance (from Jan 2021 to April 2022). This work leveraged the established CDC-funded VISION network of healthcare systems representing 10 US states to give them an adequately large sample size to evaluate the waning VE against moderate and severe outcomes under various circumstances.

I thought the authors used highly rigorous methods and interpreted their results appropriately, so I will not recount the many positives of this impressive work. I think the paper could be published essentially as is. I only have a few comments for the authors’ consideration that might further strengthen an already strong study.
Comment 1: In the Supplementary Figures, I noted that the outcomes don’t always correlate with SARS-CoV-2 test positivity for some of the sites (e.g., HealthPartners, Columbia University, Kaiser Permanente Northwest), which I suspect reflects differences in geographical scope (i.e., a specific sub-state catchment area for the site’s outcome data vs. state-level SARS-CoV-2 testing data). I understand that the authors intended to adjust for the “7-day average SARS-CoV-2 test positivity in the area of the encounter” but how appropriate is it to use state-level SARS-CoV-2 test positivity given the inconsistent correlation (at least for some sites)? Would it be possible to use site-level (instead of state-level) SARS-CoV-2 positivity data instead? Perhaps I have misunderstood and that is what the authors actually used. If indeed that is the case, perhaps the figures could be revised to reflect that.

Response: Thank you for letting us know this was unclear. In the VE models themselves, we included percent SARS-CoV-2 positivity as an indicator of local circulation based on data specific to the county-level geographic area of the site (the “site-region”). For the epi-curve figures, we depicted state-level SARS-CoV-2 positivity rather than positivity at the site-region level. We have relabeled and footnoted the figures to make this clear.

Comment 2: Since the authors specified periods of variant predominance based on when a variant account for ≥50% of sequenced isolates for each site, I wonder if a sensitivity analysis using a higher threshold (e.g., ≥90%) to specify periods of variant predominance would be worth considering. I understand that doing so will reduce precision, but may increase validity in the variant-specific estimates of VE.

Response: We appreciate the desire for potentially more variant-specific VE estimates. However, the rise of the Omicron variant was so rapid that raising the predominance threshold to say ≥75% would change the Omicron period by only an average of 4 days and would be unlikely to substantially influence results. Furthermore, repeating the analyses with a different threshold would substantially delay the revision.

Comment 3: A recent paper by Doll et al (https://pubmed.ncbi.nlm.nih.gov/35325923/) suggested that inclusion of influenza cases among test-negative controls can lead to downward bias of COVID-19 VE estimates due to correlated influenza and COVID-19 vaccination behaviours. I noted that the list of COVID-19-like illness conditions in Table S1 included a number of influenza-related diagnoses. What proportion of test-negative controls had an ICD-9/10 code for influenza, and did this vary over time such that this bias could have affected the apparent waning of COVID-19 VE observed in this study? Perhaps it biased all the estimates, but the impact of the bias would appear larger for VE estimates that are further from the null, leading to accentuation of the impact of waning.

Response: Thank you for this comment. We read with interest the Doll analysis and its implications for bias arising from inclusion of flu-positive controls. As Doll et al. report, this bias is more likely to be problematic as the proportion of the controls who are flu-positive increases and as the VE of the alternative vaccine (i.e., flu vaccine) increases. They found that little bias arises in estimates of COVID VE if <25% of controls are positive for flu. There was a historically low (essentially nonexistent) flu season in
2020-2021; the 2021-22 flu season was mild and roughly corresponded in time with the Omicron era. In our data, during the Omicron period, about 4.5% of ED/UC controls and 1.6% of inpatient controls were identified as flu positive by laboratory testing. These percentages may be underestimates if not all flu cases were tested. However, these percentages, even if doubled, fall well below the threshold of 25%. Furthermore, the VE of flu vaccine in 2021-22 was low, about 35% or lower among adults. Thus, the bias arising from use of flu positive controls is likely minimal. To further examine this question, we excluded the flu positive controls and re-ran our models to estimate COVID VE during the Omicron era. When excluding these flu positive observations from the analysis, the VE estimates were unchanged (<2 percentage points different). For these reasons, we do not expect the use of flu positive controls to contribute significant bias in this analysis. However, we have noted the theoretic possibility in the Discussion.

Reviewer: 2

Summary:
This study investigates the waning of mRNA vaccine effectiveness against moderate (emergency department/urgent care clinic visit) and severe (hospital admission) COVID-19 in a test-negative design study. The authors have presented their study well. In particular, the authors address lots of important angles through appropriate subgroup analyses (i.e., age, immunocompromised status, type of mRNA vaccine) and sensitivity analyses. The authors find that effectiveness of mRNA vaccines wanes in the months following both second and third vaccination and conclude that these findings support current recommendations for broad use of additional vaccine doses. The major limitation of the paper is the potential for selection bias through the test-negative study design.

Importance:
The COVID-19 pandemic is ongoing, and this study adds to the body of work demonstrating the effectiveness of mRNA vaccines and the utility of “booster” vaccines, even more so for high-risk groups such as immunocompromised patients. This study will therefore directly contribute to COVID-19 vaccination policy decisions.

Originality:
As mentioned in the introduction and the discussion (p.11, ll. 283-289) of the manuscript, there is already evidence on vaccine effectiveness, and whether it wanes. This study replicates these findings in a large and multi-site cohort. The authors state the originality as assessing loss of effectiveness amongst all vaccine recipients while additionally performing subgroup analyses by age, immunocompromised status, and vaccine product. Whilst encouraging to see, the replication of findings from multiple studies does question the originality of the study.

Response: To improve the originality of this paper, we have added analysis of protection of the 4th mRNA vaccine dose, including preliminary estimates of whether the 4th dose also exhibits waning of effectiveness.

Major comments:
Comment 1: General / Design: The test-negative study design poses a risk for selection bias (e.g., due to conditioning on having a test [see Li et al, Comparison of the test-negative design and the cohort design with explicit target trial emulation for evaluating Covid-19 vaccine effectiveness. SER Conference. June 14-17, 2022] and due to only including patients in the hospital/ED/UC setting). The least that the
authors could do is to discuss this limitation. Ideally, a target trial design could be used for such a causal question (as done for example by Dickerman et al. [2022], [DOI: 10.1056/NEJMoa2115463]).

**Response:** We agree that studies using a test-negative design are at risk to selection bias. We have added the following to the limitations paragraph in Discussion:

“Second, while our test-negative design is intended to avoid selection bias from healthcare seeking behavior, it may also induce selection bias arising from factors associated with a COVID-like illness but not with COVID. For instance, use of influenza-positive individuals as controls could underestimate VE due to correlation between COVID vaccination and influenza vaccination. Because <5% of controls in our study were influenza-positive, we expect this bias to be minimal. Also, we cannot rule out selection bias arising from reliance on clinician-directed testing, although we note that virtually all patients hospitalized with CLI were tested for SARS-CoV-2.”

We also agree about the advantages of a cohort design with explicit emulation of a target RCT. However, given our focus on waning, we would emulate an RCT that differs in an important way from the RCT emulated by Dickerman, Li, and Hernan. We would emulate a challenge RCT that randomizes people to be vaccinated (or boosted) on dates that range over a number of months so that--at any later date -- we can compare recent-vs-remote vaccinees who are challenged by the same variants at the same time. The usual RCT randomizes vaccination status at T0 – this renders waning estimates more problematic insofar as they are derived from analyses of “differences-in-differences” (or ratios of ratios) that compare vaccinated versus unvaccinated individuals at an earlier versus a later calendar time (when epidemic intensity and variants may differ). We note this because even though we do not directly emulate any RCT, our methods for examining waning do have this feature – which is advantageous for disentangling waning from the ebb and flow of variants – of a kind of challenge RCT rather than the more familiar RCT.

**Comment 2:** Introduction: The Introduction section could be a bit more elaborate, for example about the previous research done on this topic and a clearer description of the primary and secondary objectives. While we read the manuscript, we felt that “waning” relates to the difference in vaccine coverage, while the manuscript mainly focuses on unadjusted repeated vaccine effectiveness measurements.

**Response:** Although constrained by word count, the language of the introduction has been modified to state the study’s objectives more explicitly, including evaluation of the effectiveness of a fourth mRNA vaccine dose. The language indicates that “waning” refers to declines in how well-protected a vaccinated individual remains as more and more time elapses after vaccination (it does not refer to changes in the percentage of the population that is covered by vaccination). Description of previous research and comparison to other studies are included in the Discussion.

**Comment 3:** It currently is also unclear why the authors only present and discuss the unadjusted vaccine effectiveness results in the main text (as well as in the abstract), while adjustment for different
subgroups seems to be the main goal. Would it be possible to explain this and to provide information about the adjusted estimates as well?

**Response:** As described in the Methods, the estimates of vaccine effectiveness (VE) for the overall adult population (including all of the subgroups) are adjusted for age and immunocompromise – the factors that define the subgroups examined in stratified subgroup-specific analyses – and are also adjusted for many other factors including calendar time, geographic area, race, ethnicity, local viral circulation, presence of comorbidities, and propensity to be vaccinated.

We explain in Methods how our VE estimates were adjusted to minimize bias from potential confounders. We also explain in Methods that we also conducted subgroup-specific analyses to examine whether VE differed by age group or immune compromise status. A subsection in Results specifically reports VE in subgroups. We have modified the abstract to make more prominent that vaccine effectiveness waned in all subgroups, but word limits preclude additional detail. VE estimates for all subgroups and time periods are provided in the Supplemental Tables.

**Comment 4:** Methods: Please discuss the selection of covariates (ideally with a DAG or similar) and why some potentially influential confounders are not captured by your study (e.g., BMI, certain occupations, etc.).

**Response:** Thank you for this comment. We have added this additional information about covariate selection to the Methods, which refer readers to the supplement that describes the derivation and selection of variables for the propensity-to-be vaccinated score in detail.

“Covariates included in the models were those determined through bivariate analyses to be significantly associated with both the outcome and exposure, as well as those specified *a priori* as established confounders, including age, race, ethnicity, presence of respiratory and non-respiratory comorbidities, immunocompromise status, and local viral circulation.”

Some potentially influential confounders such as “certain occupations” were not examined because they are not systematically available in our database.

We appreciate the DAG suggestion and often find DAGs to be useful. However, we consider a large number of potential confounders and colliders (as described in Methods), and our focus is on how the effect of vaccination, especially booster doses, changes over time. The DAG would be more complicated and less familiar than usual, and could obscure our methods rather than clarify them.

**Comment 5:** Results: The authors provide confidence intervals (CIs) for the estimates of VE, however, the primary focus of the paper is the level of waning estimated (the difference between VEs at different time points). These estimates of waning also need CIs (or another measure of uncertainty), otherwise the reader does not know how precise these estimates are, which is key to appropriately interpreting the level of waning.
Response: Thank you for this comment. We agree. We now report bootstrapped CIs for the waning estimates that we report.

Comment 6: Results: We suggest providing the CIs for every estimate of VE in the main body of the text (in addition to the tables).
Response: CIs for all VE estimates have been added.

Comment 7: Results: Would it please be possible to include forest plots for the adjusted VE analyses, too?
Response: All VE estimates presented in the forest plots are adjusted VE estimates. We have added information to the Figure captions to make this explicit. If the reviewer is referring to forest plots for VE estimates among specific subgroups of interest, these are presented in Figure 3 for the most important VE estimate, VE against hospitalization during the Omicron era. Figure limits preclude additional forest plots. However, subgroup-specific VE estimates for additional time periods and for the ED/UC outcome are presented in the Supplemental Tables.

Comment 8: Discussion: we would have liked to see a discussion about the fact that the uncertainty around the estimates of VE is particularly large for younger age groups and immunocompromised (IC) individuals.
Response: We added this sentence to the limitations described in the Discussion:

“Finally, although our sample includes enough outcome events to yield precise VE estimates for the overall adult population, estimates of VE against hospitalized COVID were less precise for young adults and immune-compromised individuals due to smaller sample sizes.”

Comment 9: Please discuss the results of the sensitivity analysis including “negative control exposure” (p. 9, l. 269, VE: 8-23%) and provide results by age subgroups, too.
Response: The “negative control exposure” examines risk of COVID during days 1-14 after dose 1 before vaccination would be expected to induce much protection. If our sensitivity analyses found that vaccine effectiveness appeared to be high during these 14 days, we would interpret this as evidence of residual confounding. However, it is reasonable to expect a small amount of vaccine-induced protection to arise during this 14-day period especially later in this period. Our sensitivity analyses yielded results that were consistent with this expectation: vaccine effectiveness estimates for this 14-day period ranged from -5 to 24%. We have added to the sentence reporting these results to note their implications:

“In the third sensitivity analysis, VE ranged from -5 to 24% among patients whose index event was <14 days after the first dose consistent with little vaccine-induced protection during this 14-day period, which is reassuring because high VE would have been evidence of residual confounding.”
This 14-day period does not include enough events to stratify by age group.

**Comment 10:** Please provide footnotes in the supplementary tables with information about which variables were adjusted for (they might be lacking due to a formatting issue).

**Response:** Footnotes have been added.

**Comment 11:** Discussion section: Whilst we admire the translation of results into absolute numbers in the Discussion section, given the previous point, we wonder if the numbers presented are valid.

**Response:** This paragraph in the Discussion is intended to help readers assess how much waning is enough waning to be relevant to clinical decisions, and policy decisions, regarding the need for boosters (and other protective measures). We illustrate how to estimate the number of additional COVID hospitalizations that would occur when there is a large COVID surge in vaccinees whose protection had waned by 20 percentage points. We have re-checked this illustrative calculation. The peak hospitalization rate used in this calculation – about 1500 per million per week in the US in January 2022 – is consistent with the monthly rates reported on the referenced CDC website. We have updated the amounts of waning in this illustration so that the calculation aligns with our updated waning estimates.

“For example, the rate of COVID-19 hospitalizations reached about 1500 per million unvaccinated adults per week in January 2022 in the U.S. [30]; if incidence surges that high again, then for every million adults who lose 20 percentage points of vaccine protection there will be about $1500 \times .20 = 300$ additional COVID-19 hospitalizations per week compared with no waning effect. During the Omicron period, VE waned within 6 months of dose 3 by about 20 percentage points among those without immunocompromise and by more than 40 percentage points among those with immunocompromise.”

**Comment 12:** Please give more reasoning on the statement in the discussion that you consider the waning “clinically significant” (p. 11, l. 325), as the final outcomes of the patients remain unclear. We suggest maybe not referring to this as “clinically significant.”

**Response:** It is sometimes said of a large study’s findings about a treatment effect that even though the treatment effect was found to be “statistically significant,” meaning that it is not due to chance or confounding, the treatment effect may not be big enough or important enough to be “clinically significant.” We believe that the prevention of a large number of COVID hospitalizations is important enough to be “clinically significant” to clinical decisions (regarding whether an individual should get a booster) and policy recommendations. Although we don’t know the “final outcomes” of the patients in our hospital sample – we don’t know how many died or now have “long COVID” – we think that SARS-CoV-2 infections requiring hospitalization are typically important to prevent, and that is the meaning of “clinically significant” that we have in mind in this sentence. To clarify this sentence, we have deleted the words “clinically significant” and revised it as follows:

“This amount of waning is enough to be relevant to clinical and policy considerations regarding the need for boosters or other protective measures.”
Minor Comments:

Comment 1: We would suggest relating the percentages in Table 1 to the respective total number of participants in each column (e.g., SARS-CoV-2 Negatives in January 2021 divided by total number of SARS-CoV-2 Negatives). This would enable easier comparison of characteristics between groups.

Response: Table 1 provides row percentages to contrast the difference in each factor across case/control status (to which the standardized mean difference refers). We chose this orientation so that readers could easily see which factors differed across case status and thus were potential confounders to be included in the propensity-to-be-vaccinated score.

Comment 2: It should furthermore be discussed that the geographic areas that were “adjusted” for were quite large considering the authors’ goal to establish comparability. A full adjustment for the waning of vaccine-induced immunity might therefore not have been possible. So, we suggest removing “clearly” from line 295 on page 10.

Response: Agreed and wording revised per suggestion.

Comment 3: We wonder why it is stated on p. 7 l. 193 that p values < 0.05 were considered statistically significant, while no p values are included thereafter or referenced to. Maybe this addition is not needed.

Response: Agreed. Reference to the p values was removed.

Comment 4: P. 5, l. 159: Ideally, please describe the rationale behind the choice of (a) categorising the time variable, (b) 2-months, and (c) the choice of subgroup categories (e.g., age groups) (p. 7, l. 182)?

Responses: (a) We categorized time-since-vaccination and age, rather than model them with polynomial terms or splines, to make it easier for readers to interpret our results. The cut-points that we used are somewhat arbitrary – we tried to select round numbers of months-since-vaccination and years of age that have been used by others, and that delineate the differences in VE that our data give us power to examine.

(b) Two-month categories of time-since-vaccination let us characterize the trajectory of the waning of vaccine effectiveness. If we used narrower categories, the data would be so sparse in each category that the category-specific VE estimates would be too imprecise to interpret.

(c) Our three broad age categories were chosen to be relevant to CDC policy considerations and to align with age categorizations used in other studies. Although a 50+ age group might have been helpful for some more recent policy considerations, we tried to stay consistent with the age categories we have used since we started monitoring vaccine effectiveness more than a year ago.

Comment 5: We have concerns about potential correlations between controls, as they could participate in the study multiple times (p. 5, l. 147, 148) and would ideally like to see this discussed.
Response: The percentage of persons in the hospital sample with more than one hospitalization was 4% during the Omicron period, 4% during Delta, and 3% pre-Delta. The percentage of persons in the ED/UC sample with more than one ED/UC encounter was 11% during Omicron, 12% during Delta, and 10% pre-Delta.

By design, in every analysis we excluded all events that occurred to a person after his (or her) first test-positive event.

Our methods are intended to yield VE estimates consistent with those that would be obtained by an RCT or cohort study that used Cox regression to examine censored time-to-event follow-up. If the target outcome of such a study was the hazard ratio for a first COVID hospitalization (after randomization or baseline), then follow-up would be censored after a first COVID hospitalization but not after a test-negative hospitalization. An ideal case control study would use risk set sampling with replacement, and so a small percentage of controls would appear in more than one matched set of cases and controls. Our test negative design is a kind of case control study in which test-negative controls can appropriately be informative more than once. We note that each conditional logistic regression model was highly stratified, by calendar week and site, and so no person appeared more than once in any stratum.

Comment 6: Could you provide some rationale why your focus was on mRNA vaccines?

Response: We focused on mRNA vaccines because our data partners are all in the United States, and these vaccines comprise the vast majority of COVID vaccines given in the United States. Also, the other vaccine used in the United States, the Janssen (Johnson and Johnson) vaccine, has a different mechanism of action. We added a statement to this effect in the Methods under Study Design.

Comment 7: If the authors can find space: to further align the abstract with the main text, suggest including 'with a focus on three-dose protection against severe disease during Omicron period' in the objectives, as set out in the introduction. Especially as this is one of the unique features of the study.

Response: Agreed. Suggested wording added to the abstract.

Comment 8: We would have preferred a flow chart (STROBE statement 13c) to show the disposition of study participants over the description of included and excluded participants in Tables S7 and S8.

Response: Agreed. We have converted the enumeration of included and excluded participants to flow charts.

Comment 9: If sample size permits, we suggest presenting the same time intervals for the IC subgroup as for the non-IC subgroup in the Figure with the forest plots for better comparability (p. 24).

Response: In general, we chose to present as many time intervals as the data would support (for each subgroup) and, unfortunately, we did not have sufficient sample size in the immunocompromised subgroup to include all of the time intervals that are represented for the non-immunocompromised subgroup.
Comment 10: Please state explicitly that the data was collected retrospectively.

Response: Added this to the Methods under Study Design.

Comment 11: In some tables (e.g., S9 and S10) there are green triangles in the upper left corner of a subset of cells, which need to be removed.

Response: Thank you. Triangles have been removed.

Comment 12: It could be discussed as a limitation that the final clinical outcomes remain unclear (p. 10, l. 291)).

Response: This study was designed to study hospitalizations for COVID. Hospitalizations for COVID are our “outcomes” regardless of whether the patient ultimately died or developed “long COVID.” The effect of vaccination on the risk of a COVID infection requiring hospitalization is important to understand. It would also be very important to study COVID mortality – this could be another study. But it is not a shortcoming of this study that mortality is outside our scope. We are concerned that a discussion in the Limitations paragraph of the fact that we didn’t examine mortality or other final clinical outcomes, would suggest to readers that our findings need to be qualified and interpreted “with a grain of salt” because of this.