Responses to reviewer comments

Recommender comments

We now have detailed reviews from 3 reviewers, who all agree that the work is timely and well designed. They have made some suggestions to improve the study and analysis plans. So I invite you to address the reviewers' comments and submit your revised manuscript, which may or may not be sent back out for review.

One reviewer advocates using only one statistical framework (i.e., either frequentist or Bayesian, but not both). I agree with the reviewer that it creates room for analytic flexibility. On the other hand, it is also encouraging when both frameworks agree on the robustness of a result. So I would recommend that you specify all the priors assumed in your Bayesian tests as the reviewer recommends, but continue to use both frameworks to report the statistical results. The other two reviews also provide some useful conceptual and design suggestions.

Dear reviewers,

Thank you for your constructive and comprehensive feedback on our Stage 1 RR manuscript. Please find below our responses to your comments.

Reviewed by Phivos Phylactou

I have carefully read the Stage-1 report by Weinerova and colleagues, who propose an investigation of the effects of Long COVID on different memory modalities and different types of stimuli. In addition, the authors propose the investigation of differences in memory performance between fully vaccinated and not fully vaccinated individuals. The potential findings of the proposed study can lead to helpful insight regarding the effects of COVID-19 and Long COVID on memory.

Below, I provide some suggestions, which I think will help strengthen the current registration and subsequently benefit the conduct of the study.

1A. The scientific validity of the research question(s).

Overall, the authors give sufficient information to support the validity of their research questions, making a strong case as to why understanding the specific aspects of memory that are affected by COVID-19 is necessary.

Thank you for the positive feedback regarding our research questions. We address the concerns you have raised point by point below.

I was, however, confused regarding the specific "type" of the current report. In detail, the authors aim to replicate a previous finding (Guo et al., 2022b) showing memory impairments after COVID-19, but also aim to extend the previous findings to investigate whether the memory impairments might be specific to a particular modality or stimuli type.

For example, on p. 3 the authors report *"[h]owever, as the tasks looking at nonverbal item memory and verbal associative memory were not included […]"* and in p. 5 *"[t]o the best of our knowledge this is a first study attempting a replication […]. We aim to replicate this result and also extend it […]."* As a reader, I felt that it was unclear whether the primary goal of the study is to replicate the previous finding or to study the specific effects of memory deficits. Seeing beyond this confusion, I understand that the authors aim to investigate the specific effects and in doing so, they will also replicate the previous finding. However, I think that the report will benefit if the focus remains to what the authors consider the primary aim of the study, given also that this is a pre-registration. In detail, if the primary aim is the replication of Guo et al. 2022b, then more details will be required to highlight the need of this replication. Alternatively, if the primary aim is studying the specific memory effects after COVID, then the replication of the previous finding can be discussed later (e.g., can be potentially used as a quality check if the authors are confident in the previous finding) or omitted from the primary aims of this Stage 1 report (e.g., and discussed in the discussion section of a Stage 2).

Thank you for this comment. As you said, our primary aim is to extend the finding by studying the specific deficits. However, as this is still a relatively new topic, we believe that first establishing the presence of the previously reported effect through replication was key to ensure the quality of the data and the reproducibility of the memory effect. We have modified the final paragraph of the introduction (p.5) to clarify this:

"In the current study we aim to extend published literature on the association between SARS-CoV-2 infection and cognition. A previous study (Guo et al., 2022b) has shown that there was an effect of infection status on memory. Our primary aim is to extend this result. We will use an improved design which allows us to replicate the previous effect obtained by Guo et al. (2022b), but to further disentangle the effect of infection status on various components of long-term memory, namely, memory type (item vs. associative) and stimulus type (verbal vs. non-verbal). To this end we will analyse data from a modified version of the Guo et al. (2022b) study, which includes verbal and nonverbal versions of both item and associative memory task, collected online from a new cohort of participants."

1B. The logic, rationale, and plausibility of the proposed hypotheses, as applicable.

In general, the 9 hypotheses proposed by the authors are well thought of and reflect the theoretical foundation of their work. Below, I share some suggestions to the authors, in hopes of improving the current report.

My first suggestion is related to my earlier comment, where I suggested considering focusing the report to either the replication or the extension of previous work. If the authors decide to take this suggestion into consideration, then the hypotheses (Hs) related to Q1 could be adjusted to focus on the primary aim of the study. For example, if this is a replication study, then the Hs of Q1 could remain focused on the replication implications of the findings. If the primary aim is the extend of this work (e.g., as in Q2), then the Hs of Q1 could serve as a positive control (i.e., ensuring that the expected memory deficits are evident, so that they can proceed to the Hs of Q2).

My second suggestion relates to the various contrasts (accuracy and RT) that the authors register. In my opinion (based on previous RR experience), using multiple contrasts to test the same theory might be problematic (increased complexity and flexibility). For example, what would the results mean if for Q1 the authors find only an Accuracy effect and for Q2 the authors find only an RT effect? This issue is also related to the authors' proposal that if RT and Accuracy show conflicting results, then the findings will be considered inconclusive. I think this can be avoided with at least one of two ways:

- **(i) The authors can consider focusing this Stage-1 RR and its hypotheses on only one primary outcome, while treating the other as a secondary (e.g., Accuracies will be used for drawing conclusions, but RTs will be analyzed during exploratory data analysis)**

- **(ii) The authors can consider using a speed-accuracy trade-off transformation as their primary outcome, which overcomes the complexity and flexibility of analyzing both Acc and RT separately to test their main hypotheses. Some resources for Speed-accuracy Trade-Off measures are provided below:**

    a. **Liesefeld, H. R., Fu, X., & Zimmer, H. D. (2015). Fast and careless or careful and slow? Apparent holistic processing in mental rotation is explained by speed-accuracy trade-offs. Journal of Experimental Psychology: Learning, Memory, and Cognition, 41(4), 1140–1151. https://doi.org/10.1037/xlm0000081**

    b. **Liesefeld, H.R., Janczyk, M. Combining speed and accuracy to control for speed-accuracy trade-offs(?). Behav Res 51, 40–60 (2019). https://doi.org/10.3758/s13428-018-1076-x**

    c. **Liesefeld, H.R., Janczyk, M. Same same but different: Subtle but consequential differences between two measures to linearly integrate**

speed and accuracy (LISAS vs. BIS). Behav Res 55, 1175–1192 (2023). https://doi.org/10.3758/s13428-022-01843-2

Thank you for raising this issue. As noted, this might be an issue for interpretation mainly when the results are conflicting (e.g. prolonged RTs but improved accuracy). Detecting these conflicts (if they arise) is potentially important for the interpretation of the data as they would show that assumptions made about the cognitive tests are perhaps more subtle. Additionally, the different patterns of impairment in accuracy and reaction times can be informative in of themselves about the nature of the deficit. Therefore, we have decided to keep both. In regards to this approach increasing flexibility, if we see effect only in one and not the other dependent variable, this will be transparently discussed.

My third suggestion relates to the overlap between Q2 and Q3, but I discuss this in the next section (1C).

My final suggestion relates to Q4. This is a very interesting question, however, considering the focus of the report, to me it seems exploratory. In other words, Q4 could have been omitted from Stage-1, without affecting the primary aims of the study. However, I congratulate the authors for registering this question and I wish to clarify that Q4 should not be removed from the report solely on the basis of this comment.

Although we do not have any specific predictions with regard to this question, it is not exploratory in a sense of the related analytical approach. That is, we have a specific plan for how these data will be analyzed. We therefore include it here, already in stage 1, as an indication that we are committed to the analytical approach that we're planning to use.

1C. The soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable).

The overall design of the study is well thought of. Some aspects of the methodology and analysis can be improved, specifically in regards to the sample size justification and planned analyses. I offer my suggestions in the sub-sections below:

Sample Size Justification:

The sampling plan and its justification are unclear and further clarifications are needed. In detail, there are no explicit details regarding the expected sample size. The authors do mention that "*the expected sample is sufficiently powered to detect the predicted memory effect, and to provide evidence for the null hypothesis*" and that "*[t]he final sample size is not yet known as data collection is currently still ongoing, but is expected to range around 450*" but this does not provide any information regarding what (if any) the initial recruiting part is. Further, in the design table, the authors state that *"[t]hese are secondary data analyses. As such […]"*. I am not sure that this is sufficient justification for

the sample size. Can the authors provide the justification of the primary study regarding the choice of the sample size? Why is the range expected to be 450? It is very important that the sampling plan is clearly described. Even if the authors have no control over this, I still think that this should be communicated clearly within the report.

As the study was planned in response to an evolving situation during the global pandemic, the sampling plan was determined based on what could be done at the time. There was no specific pre-planned sample size, as at the time the motivation was exploratory. A larger sample size was considered appropriate in order to discover important but at the time unknown and unpredictable effects, and sampling was concluded when the changing situation in the world meant that the nature of the sample was likely to change notably, given the spread of novel viral variants, different types of vaccine and a diminishing pool of control participants from which to draw.

Additionally, the authors provide findings from simulations to support that with 200 participants per group, a BF > 6 for either the null or the alternative will be evident with an 80% probability (at least). Simulations are a very helpful tool within the Bayesian framework, and they strengthen the current report. Although, more information is required to understand what these simulations inform us about. For their hypotheses, the authors will conduct ANCOVAs, which will result in various BFs for each parameter of the model (see also my comments below in *Planned Analyses*). Which BFs of their model do the simulations reflect and for which research question(s)? The authors also mention that these simulations were based on directional (one-tailed) tests. I am wondering whether these simulations were based on t-tests rather than the planned ANCOVAs. The authors can strengthen their report by providing further details regarding their simulations.

Since the initial submission of Stage 1, data preparation for the waves that will be used here was concluded, and we have received access to the full dataset. In total, 325 participants could reliably indicate whether they have or have not had Covid-19, less than what we were expecting. We therefore ran the simulations again to account for these changes with N=320 (we decreased the numbers further in case we need to exclude some of the participants; note that to avoid having any unwarranted knowledge about the data prior to acceptance of Stage 1, we have not followed the full exclusion protocol and the final sample size is based on our estimation).

With 320 participants in total, divided into the same proportions as they are now at the overall sample size (90 for No-COVID and 230 for COVID group) the power to detect the effect of Cohen's F= 0.19 is 79%. This information is provided on p.6 of the revised manuscript, which now reads:

"The effect size from the memory factor detected in Guo et al., (2022b) translates to Cohen's F of 0.19 (partial eta square = 0.03). Based on this effect size, power simulations with Bayes

Factor (BF) of 6, repeated over 1,000 iterations, indicated that with 320 in total, and with group numbers imbalance proportional to the one in our data we should be able to detect a positive (one-tailed) result in 79% of iterations (equivalent to power of 0.79). Our simulations further suggested that we should be able to detect a true null effect in 75% of the iterations. Thus, overall, the expected sample is sufficiently powered to detect the predicted memory effect, and to provide evidence for the null hypothesis."

The simulation code has been adapted from the following source: https://github.com/MRC-CBU/cbu_bayesian_sequential_designs.

With regard to the statistical approach used for the simulations, as now noted on p.7, a Bayesian t-test was used. The effect of interest is the interaction effect (COVID status X Memory type/process). This effect can nevertheless be investigated with a t-test, e.g., by subtracting performance for associative memory from item memory then using a t-test to contrast between the COVID and Non-COVID group. Note that the effect size that was simulated (Cohen's T, which is equivalent to Cohen's F=0.19) was obtained in the previous study for the effect of interest after already accounting for any covariates (or other effects) included in the model. Therefore, this simulates the probability that the current sample will reveal an effect of this magnitude, after accounting for any additional factors/covariates.

Planned Analyses

The authors appropriately decided to use ANCOVAs to test their hypotheses. As mentioned earlier, these ANCOVAs will result in multiple BFs according to the parameters of each model. The authors describe that they would infer their conclusions based on a BF > 6 for the alternative or the null, though it is not clear to which specific BF they are referring. Is the decision threshold specific to a specific parameter or interaction of the model? If so, which one? For simplicity, one suggestion would be to base the decision based on the BF of the null model (i.e., if BF > 6 is evident either in favor or against the null model), but I leave it with the authors to decide which model parameter is best fit for their conclusions.

We plan on using the BF of the full model against a model including all but the effect of interest as the main parameter for the decision threshold. We have now specified this in the analysis section on p.11 to make this clearer.

"The Bayesian analysis will be performed in JASP (JASP Team, 2023) using the uniform prior (default setting in JASP) and the default setting of coefficient priors. The Bayes Factor > 6 will be used to infer conclusions for the alternative or the null hypothesis based on the Bayes Factor of the full model against a model that includes all factors apart from the effect of interest."

An additional suggestion concerns the overlap of analyses between Q2 and Q3. I do not have strong feelings about this, but I will provide my thoughts so that the authors can decide whether or not they want to take this suggestion into consideration. Specifically, the planned analysis for Q3 contains all parameters that will be included in Q2. Therefore, it seems that the planned analysis for Q2 can be tested directly from the model proposed in Q3. Further, the two models can be compared to identify which one fits the data best (i.e., Q2 2-way vs Q3 3-way).

We are aware of the overlap. However, we feel that including them as separate research questions makes our aims clearer and our approach more explicit.

Further, clarifications are necessary regarding to the normalization that the authors propose. The authors mention that the dependent variables will be normalized, but no details are provided regarding the normalization procedure (e.g., *z*-scores?).

Thank you for pointing this out. As we now clarify on p.12, we will use z-scores, applied through the R base function scale.

- https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/scale

We will make our code for all analysis available. We have now included code availability statement on p.14.

1D. Whether the clarity and degree of methodological detail is sufficient to closely replicate the proposed study procedures and analysis pipeline and to prevent undisclosed flexibility in the procedures and analyses.

The authors provide great detail describing their methods and analyses. Further suggestions are provided below to reduce flexibility and to allow replicability of the work.

Analytic Flexibility

The authors propose the use of both frequentist and Bayesian analyses to test their hypotheses. I will argue against this, since using both approaches concurrently can lead to confusion but also provide more degrees of freedom to the researcher, and thus cause an issue of analytic flexibility (for similar arguments see these preprints: Dienes, 2023 https://doi.org/10.31234/osf.io/2dtwv & Phylactou, 2023 https://doi.org/10.31234/osf.io/dthns). I advise the authors to focus and stay within one framework of statistical inference. Additionally, if the authors decide to stick with a Bayesian approach, the priors of the coefficients for the planned ANCOVAs should be mentioned, in addition to the model priors.

Thank you for your comment. We will use the default settings in JASP for the coefficient priors. We have now added details of this to the analysis section of the RR (p.12), which now reads:

"The Bayesian analyses will be performed in JASP (JASP Team, 2023) using the uniform prior (default setting in JASP) and the default setting of coefficient priors."

We strongly believe that in the current case it is important to use both Bayesian and Frequentist statistics since the effect that we are replicating was originally found using frequentist statistics. Hopefully being able to detect the effect using both frameworks will be a good signal for the robustness of the effect. Any conflicts between the two approaches might also be of interest, and including both ensures that they are transparently reported and openly discussed, if they occur.

Replicability

To allow replicability, the authors can provide further details regarding their methods. For example, the specific details for the non-verbal and verbal memory tasks are not known (e.g., stimuli size, items, etc.). For replicability, the authors could describe in more detail the main tasks, or provide the materials in an accessible format so that future studies could access the details of the task design. In a similar vein, it would be beneficial for the report if a data accessibility statement is provided (preferably stating open access to data and material).

As this is secondary data analysis, we have limited recourse in influencing when and how the data is going to be shared. As of now there is planned release of the stimuli on gorilla.sc Open Materials and the data will eventually be shared once they have been used for additional studies planned by the group that have collected them. We have added a relevant information on the presentation of the stimuli on p.8. All information about timing is also included. As the data was collected online on individual's personal devices using Gorilla, stimuli-size will have differed between participants. Gorilla is designed to allow researchers choices as to whether stimuli sizes are fixed or adjusted based on screen size. To ensure that the tasks as a whole would function on as wide a variety of screen-sizes as possible (with the exception of preventing them being conducted on phone screens as these were deemed too small), they were made flexible.

1E. Whether the authors have considered sufficient outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses or answer the stated research question(s).

The current design is sufficient to answer the proposed research questions. The authors can decide if replicating the previous findings can serve as a positive control for their study, based on my suggestions mentioned in the previous sections.

I hope that the authors find my suggestions insightful.

Respectfully,

Phivos Phylactou

Reviewed by Mitul Mehta, 07 Dec 2023 18:14

The aims are to understand covid related cognitive impairment with the first hypothesis asking if there is a relationship between covid status and item and associative memory. The study is embedded/part of a longitudinal cohort. I think it is important to know more about this cohort and wider aims (see below on 'who' this study is about) and also consider alternative interpretations of the tasks.

Thank you for your comment. We have now updated the participant section (p.6) as follows to reflect more details on the wider aims of the cohort:

"The data is being collected by members of the Cambridge Cognition and Motivated Behaviour Lab (CambLab) as part of the Covid and Cognition study ("CovCog"). This multi-cohort longitudinal study has published early findings (Guo et al., 2022a; 2022b) with their first cohort. This work will concentrate on the new cohort. In total, 430 participants have taken part in the study. After excluding duplicates, unfinished questionnaire entries, participants unsure of their Covid status, and participants who have not completed at least the two memory tasks that our analysis mainly focuses on, there are 325 remaining participants in the sample (COVID group N=232, No-COVID group N=93). Compared to the previously published study, the new dataset includes additional tasks and measures as well as more detailed information about the vaccination status of the participants (details below).

The general aim of establishing this cohort was to study the effect of Covid-19 on cognitive function in adults. The term Long Covid is in connection with the sample referring to those who had confirmed Covid-19 diagnosis in the past and are experiencing lasting symptoms. A medical diagnosis of Long Covid was not a requirement to take part in the study. Compared to the previously published study, the new dataset includes additional tasks and measures as well as more detailed information about the vaccination status of the participants (details below).

Participants were recruited through word of mouth, social media platforms such as Long COVID Facebook support groups, from Addenbrooke's Hospital Long Covid clinic, and the

Prolific recruitment site (https://www.prolific.co/) through majority English-speaking countries (UK, US, Ireland, Canada, Australia, New Zealand, South Africa). Recruitment ran between February 2022 and May 2023."

Alternative interpretations of the tasks and results will be considered in Stage 2.

The main aim of the work is good as there is a known vulnerability of associative memory to impairment relative to item memory across multiple conditions. While it is reasonable to ask if the same is the case for covid-related cognitive deficits, no reason is given as to why it is expected to be so after COVID, except that this is a common pattern of deficits. What would be the reasoning for COVID to produce cognitive deficit patterns similar to other conditions? Is there evidence of damage, or dysfunction in the relevant brain networks for example? Some more information would be very useful here.

Thank you for pointing this out. We have now added following clarification on p.4.

"Indeed, studies using histological techniques in brains of deceased Covid-19 patients (Bayat et al., 2022) and animal models (Klein et al., 2021) found reduced neurogenesis in the hippocampus. Hence, given this relative greater vulnerability of associative memory in various circumstances, and findings of reduced neurogenesis in the hippocampus caused by Covid-19, we predict that mnemonic deficits caused by Long Covid would have greater effect on associative (vs. item) memory."

This is an observational study of the deficits experienced by patients. In addition, the impact of vaccination status will be assessed. The authors do mention long-covid regularly and the recruitment method includes long-covid groups. There is no formal, internationally recognised definition of long-covid as far as I know, and no criteria are given in the manuscript, making recruitment based on long-covid more difficult. Clear recruitment criteria around the long-term symptoms are required. There is a risk of self-selection among those who are informed of the study towards those participants with cognitive difficulties. The current recruitment routes and methods therefore allows for different inferences compared to simply recruiting on the basis of previous infection and vaccination without selecting. I urge the authors to reflect on precisely who their research questions are about and about what they would like to make inferences (e.g. long-covid, SaRS-CoV2 infection).

Thank you, that is indeed an important consideration.

Here, we use the term Long Covid to refer to those who had confirmed covid in the past and are experiencing lasting symptoms. Because they're not in the acute phase of the disease anymore, (cognitive) effects are attributed to long-lasting effects of previously being ill (i.e. Long Covid in short). We are not using diagnostic criteria of Long Covid, but there are long-lasting effects of Covid and here we aim to identify the cognitive ones. We have added clarification on p.6 (copied in our previous response above).

Please also consider including questionnaires on other potential important factors such as depression symptoms and trait anxiety levels and consider inclusion of these as covariates for the group comparisons or correlates within the covid group.

Unfortunately, as this is secondary data analysis, we are unable to include any additional measures. The potential effect of those on our results will be discussed in the Discussion section of our Stage 2 manuscript.

Also, in our previous study we found a big difference in cognitive impairment between those with confirmed and suspected COVID. Given other infections do exist I would urge the authors to focus their primary comparisons on the confirmed group.

Thank you for highlighting this important finding. We have previously been unsure as to how many participants would indicate that they have either had Covid-19 based on experience symptoms alone or that they don't think they have had Covid-19 but have experienced some symptoms. Now that data collection has ceased we know there are 51 in the former case and 17 in the latter. Based on the concern you have raised we have excluded those who have indicated experiencing symptoms but do not think they had Covid-19 (N=17). In regards to the 51, 28 out of those have indicated they have Clinical Long Covid Diagnosis, further 4 have been diagnosed by GP or health consultant based on symptoms and 1 has had positive antibody test. Therefore, we will exclude the remaining 18 who have indicated Covid solely based on their symptoms or symptoms of close contacts. This has now been specified on p.11.

"For the purpose of the analysis, participants' Covid status will be established based on their answer to the question "Q3.01 Have you had Covid-19?" (see appendix A for exact answer). Participants unsure of their Covid status will be excluded from the analysis ("unsure" is defined as answering "Yes" based solely on symptoms or symptoms of close contacts, or answering "No" but reporting experiencing symptoms). Otherwise, if they answer "Yes with positive PCR test", "Yes with positive Lateral Flow Test", or "Yes, no test", they will be assigned to the COVID group for analysis. If they answer any of the "No" answers, they will be grouped in the No-COVID group."

I suggest for page 4 of the document our own paper is included as a reference if the authors want to cite findings from covid infection in general (Hampshire et al 2021 – already cited elsewhere). Of course, do check if relevant as I do not insist our paper is further cited, but as our largest effect size was in word finding this aligns very well with your point, but outside of a long-covid group.

We have now added the citation in a relevant place p.4.

"This distinction, however, is particularly important as verbal difficulties have been reported among Long Covid symptoms (Miskowiak et al., 2022) and have been detected in people who have recovered from Covid-19 (Hampshire et al., 2021).“

For the tasks I have two queries/concerns and suggest further consideration or justification is given.

First for the verbal memory task on page 7/8, how can the researchers be sure verbal mediation strategies are not used, making the non-verbal task more like a verbal task? Also, is there a concern that those who take longer will be tested (on average) at a later time compared to those who respond faster? Could this have knock-on effects for the associative recognition task.

Second, the WCST is a very old task and problems with this have been expressed in the literature for a number of decades. Some of the problems with scoring are highlighted here: https://doi.org/10.3758/s13428-021-01551-3.

Other issues are conceptual and exemplified in the Id/ed literature (e.g. Downes et al (1989). Impaired extra-dimensional shift performance in medicated and unmedicated Parkinson's disease: Evidence for a specific attentional dysfunction. Neuropsychologia,27,1329±1344.)

Thank you for raising these points. These are all valid concerns for the interpretation of the results (once obtained). However, as we are using secondary data that have already been collected, there is little we can do to address them. In Stage 2, we will discuss these in-depth, in the context of the results that are obtained here.

The analyses seem appropriate for the data.

Reviewed by Dipanjan Ray

I am pleased to provide my review of the registered report submitted for consideration. Overall, this report presents a well-executed study, marked by

several strengths and contributions to the field. While the report has several noteworthy merits, there are also a few caveats that warrant careful consideration and potential refinement. In the following review, I will discuss both the strengths and limitations of the report, offering constructive feedback to help enhance its overall quality and impact.

The research question is scientifically valid, addressing a timely and relevant issue. The study seeks to explore (i) the impact of SARS-CoV-2 infection on memory function, (ii) potential variations in this effect concerning memory type (item vs. associative) and stimulus type (verbal vs. pictorial), and (iii) whether these effects are moderated by vaccination status. These inquiries are firmly rooted in current scientific concerns and draw from an existing body of evidence. The hypotheses put forth in the report are logical and plausible, with clear connections to existing literature. The distinctions between item and associative memory, as well as verbal and pictorial stimuli, are well-founded in the realms of cognitive psychology and neuroscience research. The hypotheses are articulated precisely and flow directly from the research questions.

The design is straightforward, the analysis plans seemed appropriate. The statistical power analysis supports the feasibility of the sampling plan. The inclusion of Bayesian ANCOVA with uniform priors is particularly noteworthy for its transparency and ability to quantify evidence for the null hypothesis. The methodological information provided is adequate for replication, offering clear group definitions as well as concise descriptions of cognitive tasks and analysis plans.

My primary concern revolves around participants potentially encountering difficulties when accurately recalling specific details. These details encompass the timing of their last vaccine dose, the type of vaccine received for each dose, whether they tested positive in a Rapid (Lateral Flow) or PCR test, and the timing of any confirmed COVID-19 diagnosis in relation to their vaccination history. This task becomes even more challenging considering that COVID-19 has been demonstrated to impact memory function, which is the focus of the current study.
To address this issue, I propose two strategies, either of which the authors can consider implementing to partially mitigate this challenge:

1. Confidence Rating: Alongside each question, incorporate a confidence scale. After participants provide their response, ask them to rate their confidence in their answer using a scale, (such as 1 to 5 or 1 to 7), with 1 representing ”not confident at all” and the highest number indicating ”very confident.” This

approach allows for the identification of responses where participants may lack confidence, highlighting potential areas of uncertainty.

2. "I Don't Know" Option: Introduce a response option that allows participants to select "I don't know" or "unsure". This enables participants to acknowledge when they are uncertain about an answer rather than making guesses, enhancing the accuracy of the data collected.
Nevertheless, it appears that the majority of the data has already been gathered, with 421 out of the expected 450 subjects collected. Given this situation, there may be limited room for revisions, and I am uncertain if this goes against the original intent of a preregistered report. I would like to request clarification from the he recommender.

If feasible, I suggest including a "confidence rating" or "I don't know" options for new subjects. This could provide a preliminary understanding of how the aforementioned issue might impact the interpretation.
Furthermore, the absence of inquiries about participants' memory competence prior to the COVID-19 pandemic represents a potential limitation in this study. This information would have served as a valuable reference point.

Without this baseline data, it becomes difficult to ascertain whether other preexisting factors that cannot be captured based solely on age, education level, or medical history might be influencing any detected variations in memory performance.

Thank you for your positive feedback on our manuscript and for raising these important suggestions, which all address valid concerns regarding potential interpretation of the results (once obtained). Unfortunately, however, this study entails secondary data analysis, and data collection had already finished. As such, we are unable to implement any changes to the design of the study. Nevertheless, in Stage 2, these caveats will be mentioned, and we will thoroughly discuss how these issues might constrain the interpretation of the results.

In summary, this registered report presents a well-structured study with clear and replicable procedures. It effectively addresses a significant research question, supported by a logical and well-founded hypothesis. The analysis pipeline is robust, incorporating appropriate statistical considerations. Nevertheless, it is important to acknowledge certain caveats, such as the absence of a test for the reliability of participants' memory or the lack of baseline memory competence data, as previously mentioned.