**Changes in memory function in adults following SARS-CoV-2 infection: findings from the Covid and Cognition online study**

*Josefina Weinerova, Sabine Yeung, Panyuan Guo, Alice Yau, Connor Horne, Molly Ghinn, Lyn Curtis, Francess Adlard, Vidita Bhagat, Seraphina Zhang, Muzaffer Kaser, Mirjana Bozic, Denis Schluppeck, Andrew Reid, Roni Tibon, Lucy Cheke*

Stage 1 Review for PCI-RR - ArticleID #547

*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.

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).

**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:

1. 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)
2. 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:
   1. 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
   2. 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>
   3. 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>

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.

**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.

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.

**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.

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).

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?).

**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.

**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).

**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