Risk and compliance over time - Response to second PCI RR peer review

Dear Editor,

We would like to thank you for taking the time to handle the revision process of our stage 1 manuscript submission. We are happy to see that the reviewers were satisfied with our prior revision, and we found that their suggestions on how to improve the manuscript further to be clear and well-informed. We have revised the manuscript and R script based on this input, and we believe that this process has improved the quality of the submission.

Please find our responses below and the action we have taken based on each issue raised by the reviewers. We hope that the editor and reviewers agree that these revisions have improved scientific rigor of the study.

Best wishes on behalf of the authors,

Sebastian B. Bjørkheim

Editor:

Dear authors,

Thank you for the in-depth revision of your manuscript. I received reviews from the three reviewers, and they all were satisfied with the revision of the manuscript. One reviewer provided several minor suggestions to implement for the next round.

Two reviewers took time to review the R script and had several questions and suggestions to improve it. These points need to be addressed before acceptance of the Registered Report.

Regarding the issue of the result section with reviewer 2, it is actually possible to write a result section with dummy results to improve understandability of the procedure for a RR. It is up to you to do so, but if you don't, please ensure that all points regarding the R script and the procedure are sufficiently detailed in the manuscript/code and response to the reviewers in the next round.

Best regards,

Adrien Fillon

Response to Editor:

Thank you for the clarification the clarification regarding the results section. After careful consideration however, we have opted to not include a results section based on simulated data in this manuscript. Instead, we have made significant improvements to the R script to ensure it clearly demonstrates how the results will be generated and presented with synthetic data. We have also added a file with tables summarizing the multiverse analysis (with dummy results),

which should further clarify the potential outcomes (see the file: Summary_of_multiverse_analysis_dummy_results on https://osf.io/5k7qw/).

We believe these revisions will make the research procedure more transparent and easier to follow. We hope these changes address the concerns raised and provide sufficient detail about the prospective results.

Peer Reviewer #1

I read the revised version of the manuscript with interest. The authors have done a great job in responding to the comments of the editor and reviewers. The reformulation of the hypotheses, the planned multiverse analyses, and the steps taken to deal with the risk of bias improve the manuscript considerably.

I do, however, have a reservation about the script. I am not familiar with doing RI-CLPM with R (but more on MPlus), but the script seems somewhat different from what I am used to. In particular, it doesn't include the between components, allowing to distinguish between variance from within variance. But I am relying on Hamaker et al.(2015) and Mulder & Hamaker (2021). Perhaps the authors rely on other references? If so, it may be interesting to cite them.

References

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. Psychological Methods, 20(1), 102–116. https://doi.org/10.1037/a0038889

Mulder, J. D., & Hamaker, E. L. (2021). Three Extensions of the Random Intercept Cross-Lagged Panel Model. Structural Equation Modeling: A Multidisciplinary Journal, 28(4), 638–648. https://doi.org/10.1080/10705511.2020.1784738

Response to Peer Reviewer #1

Thank you for pointing this out! We appreciate your careful review of the submission in general and the R script in particular for this round.

We can confirm that we rely on the same sources that you mentioned, and we have now referenced our approach to that of Mulder & Hamaker (2021) in the analysis plan.

We acknowledge that the initial version of the script did not integrate the between-subjects components (i.e., the random intercepts) into the model. We have revised the script to reflect the between-subject and within-subject variance components in the RI-CLPM. After making this adjustment we believe the revised script now addresses these aspects of the model in a clear way.

Peer Reviewer #2

I want to thank the authors for their responses to my previous comments and for the revisions. I appreciate the effort put into improving the introduction, and I'm glad to see the additional

measures aimed at reducing the risk of bias. I have some minor comments that I hope will be of help, listed in no particular order:

Peer reviewer 2, issue 1:

- On page 2 the authors have added a new paragraph that I think has improved the introduction of this paper but has somewhat disrupted the flow. My comment refers to the part where they write "However, the motivation to comply with infection control measures may only partly be driven by people's self-interest in safeguarding their own health and partly driven by the desire to help or protect others (Aydinli et al., 2014) […]" The second sentence doesn't flow from the preceding sentence, where the authors emphasize that previous work has been cross-sectional. The paragraph where they do elaborate on this point now comes after the new paragraph. You could either move this sentence to show up later in the text or mention the point about studies being cross-sectional after the new paragraph, where you go into this point in more detail.

Response to peer reviewer 2, issue 1:

Thank you for pointing this out. We agree that the sentence disrupted the discussion of perceived risk as a predictor of protective behavior, and we have moved it to the preceding paragraph where we discuss the potential limitations of cross-sectional data.

Peer reviewer 2, issue 2:

- Regarding a Results section based on simulated data, as per my comment in the previous round: The authors responded that per the PCI RR guidelines, the Stage 1 manuscript should not include a Results section based on simulated data. But I couldn't find this in the guidelines, and I have seen other PCI RR Stage 1 manuscripts do this. Maybe I missed something or perhaps the guidelines have changed, but I don't think it's correct that Stage 1 manuscripts are not supposed to include this. I'm not suggesting that the authors do this for this manuscript, but this is just a general comment for future studies—I find it very helpful, not only as a reviewer but also as an author (I always end up finding new issues that I would have missed otherwise).

Response to peer reviewer 2, issue 2:

Thank you for your feedback on this issue. We initially interpreted the guidelines to mean that the Stage 1 manuscript should not include such a section. However, as you and the editor have pointed out, including results based on simulated data is acceptable and can be beneficial. While we have chosen not to include this in the current manuscript, we have made several updates to the R script to enhance clarity and demonstrate how the results will be presented. Additionally, we have included a table summarizing the multiverse analysis, which should help clarify the potential outcomes (see the file: Summary_of_multiverse_analysis_dummy_results on https://osf.io/5k7qw/).

Peer reviewer 2, issue 3:

- Very minor comment, but R syntax seems to be missing some tiny details (or maybe I missed them):
    - I have no experience with multiverse analysis but my impression is that while the code specifies different combinations for the perceived risk variable, there's no code that actually executes the multiverse analysis(?).
    - The authors write "We will compare the results of the complete case sample with different ways of handling missing data (both listwise deletion and pairwise deletion)." The R script does not seem to include code for handling missing data.
    - Related to the previous point: Would it be a good idea to specify how you will compare the results (i.e., complete cases model vs missing data models)? For instance, I imagine you'd compare the results based on p-values and coefficients but most likely p-values won't change by much given the large sample size(?) So will you then focus on the size of the coefficients instead? What will count as "different"? For instance, if a coefficient is 0.03 in one model but then 0.04 in another model, does that count as similar or different? Will you test whether coefficients are statistically different?

Response to peer reviewer 2, issue 3:

Thank you for your insightful comments. To your first point, you are correct that while the code specifies different combinations for the perceived risk variable, there is no code that directly automates the multiverse analysis. We encountered challenges in developing code that would automatically account for the time-specific indices corresponding to each of the four rounds of data collection. Despite considerable searching, we were unable to find an existing code that worked for a multiverse analysis of RI-CLPM data. We think that this probably suggests that this (multiverse analysis of RI-CLPM) may be an area with limited prior work.

We have thus only included the code that details how to manually enter the time-sensitive indices for each of the 15 multiverses. While it is not ideal, it provides a viable solution for conducting the analysis across the different operationalizations of perceived risk. We believe this is the most practical solution given the current state of the available resources.

To your second and third point, we have added the code for handling missing data, allowing for comparisons using listwise and pairwise deletion methods. However, as you correctly pointed out, systematically comparing all coefficients across the multiverse analysis (with its 15 different combinations of perceived risk items across nine hypotheses) will be a tedious process and could become difficult to assess comprehensively.

Given the complexity and the sheer number of coefficients, our approach will be to focus on reporting the general tendencies observed across the various models. Specifically, we will highlight if patterns on the key relationships between perceived risk and compliance variables change direction (or remain consistent) across the different combinations and missing data handling methods. Note that the confirmatory hypotheses are tested in the main model and the multiverse will critically inform discussion of the results. We added this to section 2.4 of the manuscript to further clarify out approach:

"We will test the hypotheses listed above by indexing the four perceived risk items and use a multiverse analysis to assess the robustness of the findings."

We will follow the example of Steegen and colleagues (2016) when reporting the results of the multiverse analysis. To ensure transparency of the results, we will make the detailed output of these analyses available in the appendix of the manuscript and on the OSF site (https://osf.io/5k7qw/) of this submission. This way, interested readers and reviewers can examine the full set of results, including all coefficients and p-values, to independently assess the robustness of our findings, while the main body of the paper will discuss the tendencies more broadly. We think that this approach will strike the best balance between the need for thoroughness with the practicality of reporting results in a way that is accessible and meaningful. We will thus not perform additional equivalence testing on the difference between coefficients, but the potential differences will inform the discussion section of the manuscript.

Reference:

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. Perspectives on Psychological Science, 11(5), 702–712. https://doi.org/10.1177/1745691616658637

Peer reviewer 2, issue 4:

- You have removed factor analysis from the Method section but not from Table 3.

Response to peer reviewer 2, issue 4:

Thank you for pointing this out! We have edited Table 3 to reflect the change in approach from factor analysis to multiverse analysis.

Peer reviewer 2, issue 5:

- I took a quick look at Orth et al (2022) and they seem to define 0.03 (for both CLPM and RI-CLPM) as a small effect(?). See the last paragraph in section "Effect Size Conventions Suggested by the Present Research". If my understanding is correct, then please revise.

Response to peer reviewer 2, issue 5:

Thank you for detecting this discrepancy! We have edited Table 3 to reflect that we consider effects equal to or above 0.03 to be meaningful in the predicted direction.

Peer reviewer 2, issue 6:

- Are you planning to use unstandardized or standardized coefficients? Please specify. This choice can impact the interpretation of the meaningfulness of associations (an association might seem either smaller or larger than the smallest effect size of interest depending on whether it's standardized or not). Also please make sure the R code reflects this choice.

Response to peer reviewer 2, issue 6:

We will report standardized coefficients as this better reflects the relative strength of relationships between the variables in the model.

We have specified this in the summary function in the R script by including "standardized = T" in the model execution.

Peer review #3

My main issue with the original manuscript was that there was potential to better control for bias given the pre-existing data and the fact that part of it has already been analyzed. I find the countermeasures suggested by the authors in the revision to address these concerns.

Thus, I now only have two minor comments, which are only meant as suggestions for improvements.

Peer reviewer 3, issue 1:

- Paragraph that runs from p 2 to p 3 (starting with "The psychological aspects of risk perception…" and ending with "…compliance to infection control measures.") is quite long and dense, and I struggle to pick up the main message. Could this be split in two, and perhaps a summarizing sentence added?

Response to peer reviewer 3, issue 1:

Thank you for pointing this out. We agree that the paragraph was long and difficult to follow. We have now split the text into three paragraphs, with one about the research on perceived risk in general, one about perceived risk in a pandemic setting and how it may be associated with compliance, and one about what other factors than perceived risk may influence compliance. We have also shortened several of the sentences, and we think that these changes have made the main messages clearer.

Peer reviewer 3, issue 2:

- Power analysis: I suggested looking into the possibility of running a power analysis. The authors reply that this requires making a lot of assumptions, which makes them skeptical of running it. I am sympathetic to this reasoning. Nevertheless, in their reply letter they also write that "*Based on the example given in that paper, we can conclude that for a cross lagged panel model with 4 measurement rounds, a sample of 1800 is sufficient to reliably detect "small" cross lagged effects of .10, at a power of .80, even with a very high degree of between-unit variance. We expect a panel sample of n ~ 2000 in our study, and we should thus expect to have a power of more than .80 to detect "small" cross lagged effects.*". I found this quite informative, even if it is just a rough estimate. Perhaps it is self-evident that such a large sample gives enough power to detect even small effects, but I think it could be valuable to spell this out for readers. Note that this would not be a post-hoc power analysis, but a sensitivity power analysis giving a rough estimate of what kind of effects one would have for example

80% power to detect given the sample size and alpha (and other relevant parameters). This is not a big point, but at least something to consider.

Response to peer reviewer 3, issue 1:

We agree that it could be added to the manuscript, and we have stated it more explicitly under section 2.1 Participants:

"Following Mulder (2023), a sample of 1800 is sufficient to reliably detect "small" cross lagged effects of .10, at a power of .80 across four measurement rounds, even with a very high degree of between-unit variance."