

Dear Editor,

We would like to thank you for taking the time to find reviewers and to handle the stage 1 manuscript submission process. All the reviewers made a number of comments and suggestions, all of which we found to be well-considered and thoughtful. We have revised the manuscript based on this input, and we believe that this process has greatly improved the quality of the submission.

Please find below a numbered list of all the issues raised in the peer review, our response, and the action we have taken based on each issue. We hope that the editor and reviewers agree that these revisions have improved the clarity, scientific rigor, and the potential contribution of the study.

Best wishes on behalf of the authors,

Sebastian B. Bjørkheim

Editor:

Dear Authors,

I now have received three deep and elaborated reviews concerning your manuscripts. This is a difficult review to handle given its high risk of bias due to the use of existing data, for which you already published studies on related hypotheses. The three reviewers agreed on four key points that prevent a recommendation yet:

First issue: H1a, and rationale for the hypotheses

The first hypothesis, especially H1a has already been tested in other published studies, and you need to detail more how this test is new, in terms of key measures. More precisely, how the operationalization differs from last studies and how this hypothesis can inform the theory beyond what was already conducted on similar measures with the same dataset. The rationale for hypothesis 3 is not sufficiently documented (Erik Løhre also mentioned that hypothesis 2 could be also more explained in the introduction).

Response:

We agree that an analysis close to what was represented in the manuscript as H1a has already been tested in this dataset, at least conceptually. We have described the findings from the previous study more in detail under section 2.3:

“Firstly, response distribution for perceived risk and compliance at T1 has been examined and reported (Sætrevik, 2021) and we have previously described the cross sectional association between perceived risk and compliance at T1 (Sætrevik & Bjørkheim, 2022). Contrary to our hypothesis, we found that those who perceive a higher risk are actually less likely to comply with safety measures. When we specifically looked at the risk of getting infected, we found a very small negative effect on compliance. However, when we asked about the perceived risk of infection for the general population, the effect on compliance was small and positive. Overall, our findings only weakly support the idea that seeing the pandemic as a threat leads to following safety measures, and only when considering the risk to the general population, not to oneself (Sætrevik & Bjørkheim, 2022). This association will not be considered among confirmatory hypotheses in the current study.”

Consequently, we have decided to remove the cross-sectional association (H1a) as a confirmatory hypothesis to be tested in this study. It's worth noting that what was formerly labeled as "H1b" has now been relabeled as "H1a" in the revised manuscript.

We suggest that the relationship between risk and compliance at time point 1 (what was previously H1a), will still be conducted under our analysis plan, but it will not be tested and reported as a confirmatory hypothesis. We propose this approach to maintain consistency in methodology across different time points and to facilitate comparison between various hypotheses (H1a-c).

Since the previous analysis used somewhat different operationalization of the variables, we believe it is preferable to rerun this analysis in the same model. To make it clear that we have renamed the former H1a to R1a in the manuscript to avoid confusion.

We acknowledge the need for further description of the rationale behind hypothesis 3, as well as the possibility of elaborating more on hypothesis 2 in the introduction, as suggested by Erik Løhre. Regarding hypothesis 3, we agree that it would be stronger if we predicted the direction of the relationship as pointed out by reviewer 1 and discussed by reviewer 3. In response to this, we have changed the last paragraph about "compliance to infection control measures" to include a discussion of why the relationship between "compliance" and "perceived risk" may change during a pandemic. The paragraph now reads:

"As the pandemic progressed, perceptions of risks related to COVID-19 may have changed for some individuals. Some may have become complacent about the situation, while others may have become more fearful as the number of cases rose and new variants of the virus emerged. Through the first year of the pandemic, new information and guidelines were periodically released, and this may have affected people's confidence in and motivation for complying with the infection control measures. On the other hand, people may have gained confidence in the control measures as it may have prevented them from getting infected with COVID-19, and they may have ascribed this to the effectiveness of precautionary behavior. As a consequence, compliance with infection control measures may contribute to perceiving less risk of getting infected with COVID-19."

We also made an addition to section 1.2 (knowledge gap) to show why we are uncertain about the direction of the relationship between prior compliance and perceived risk. The addition reads as follows:

"The combination of perceived risk and compliance to protective measures are rarely studied longitudinally, and the reverse relationship, with compliance as the predictor and risk as the outcome, is even less documented in the literature. While we should expect perceived risk to predict compliance, less is known about how past engagement with protective behavior may impact how people perceive the risk scenario."

Reviewer 2 suggested that hypothesis 3 could be made to be directional. We concede that it may be more plausible to think that higher past compliance may be related to a reduction in perceived risk, However, as far as we can tell, this does not appear to substantial empirical indications of this in the literature. After reflection, we would therefore prefer to keep this hypothesis non-directional, as we believe that investigating this relationship without assuming a direction better represents the current "state of knowledge". However, if the reviewers can point us in the direction of some literature that may support a directional hypothesis, we are open to reconsider this.

Second issue: Risk of bias

Crucially, not enough has been made to prevent the risk of bias associated by the accessibility of the data. Indeed, the manuscript is a Level 1 submission (Please see section 3.6 here: https://rr.peercommunityin.org/about/full_policies#h_95790490510491613309490336); and especially "Submissions at Level 1 or 2 will usually be required to include stringent countermeasures against overfitting, such as the adoption of conservative inferential statistical thresholds, recruitment of a blinded analyst, or multiverse/specification analysis" For this specific case, I would like the authors to include the three countermeasures proposed because of the high level of potential bias involved. Note that regarding the adoption of conservative inferential statistical threshold, I am in favor of lowering the alpha level instead of conducting a power analysis mentioned by reviewers. This is motivated by the following study: <https://open.lnu.se/index.php/metapsychology/article/view/2460>.

Response:

We agree that we may need to take further steps to assure the readers that our analyses are not influenced by bias from already having had access to the datasets. We have implemented the following countermeasures:

- *We have reduced the alpha level from $p < .05$ to $p < .01$.*
- *We will perform a multiverse analysis by testing the cross lagged panel model with different ways of combining the perceived risk items into an index. This will leave us with 15 possible versions of the "perceived risk" variable (we have excluded the option where none of the items are counted), and it will enable us to compare how robust the findings are to a particular operationalization of "perceived risk". See the statistical script for a detailed description of the analysis plan.*
- *As a part of the multiverse analysis, and to better be able to assess the robustness of the findings, we will also analyze the model with different ways of handling missing data (both listwise deletion and pairwise deletion). We will compare the results of this to the main plan of running the analysis on the participants who answered all the items in all rounds (complete cases). We expect that this approach will yield a panel of $n \sim 2000$. See also our response to the fifth issue below.*
- *We suggest that the second author (SWH) will function as a blinded analyst, as he has not had previous access to the data, and has not been part of the data curation process. To ensure that the second author will indeed be blind to the analysis output, we suggest that the first author (SBB) rename the variables "perceived risk" and "compliance" with arbitrary labels ("tango" and "foxtrot") for all the time points, before transferring the dataset to the blinded analyst (second author). The blinded analyst can thus not know if the hypotheses are supported or not when performing the analysis. We concede that this does not constitute a "fully blinded" analysis, but we believe that the relabeling of variables in combination with the multiverse analysis and more stringent alpha level will effectively guard against the risk of bias brought about by the prior knowledge of data.*
- *We have added all changes to the analysis plan to Table 3 in the revised manuscript.*

Third issue: Statistical script

Two reviewers asked for the providing of a statistical script. This is especially important for two reasons: 1) It is unclear about what kind of controls could be involved in the "positive association" hypotheses, and 2) we need to understand what will be done in the multiverse/specification analyses.

Response:

We agree that a script for the statistical analyses will help to make the statistical analysis clear and transparent. We have included an R-script for the statistical script for the multiverse analysis with the inference criteria, handling of missing data (see <https://osf.io/yvz87/>). We have tested the script on a

synthetic dataset. We have commented the script in order to make it understandable for outside readers.

Fourth issue: Definition and operationalization

More than the points above, reviewer 1 also asked authors to ensure the specificity of the question: how risk is defined and operationalized. if it is for oneself (as stated in the introduction) or for others and in general (as operationalized in some questions); but also defining compliance versus preventive behaviors.

Response:

We agree that the manuscript needed to be clearer about how we define and operationalize the variables. The “perceived risk” variable in phase 1 of the manuscript consists of four items that cover different aspects of how one assesses risks associated with COVID-19 (see “2.4 Materials” in the manuscript). We suggested in the original submission of the manuscript to combine the items into an index to make a more reliable measure of the variable. We also suggested to run a factor analysis on the items to determine the best fit. As such, the operationalization of the “perceived risk” variable depended on the outcome of this process, and may include any combination of the four items that constitute the variable. However, in the revised approach that we now suggest, we will conduct a multiverse analysis that will test the hypotheses based on all combinations of the items (all 15 iterations). This will ensure that the readers will know how robust the results are to changes in the operationalization of the perceived risk variable. Note also that we have added a section in the background literature to highlight recent literature that suggests how concern for other people’s health risks may be an important predictor of compliance to infection control measures. See additions to section 1.1.1 Perceived risk.

As for the “compliance” variable, reviewer 1 raised the issue of whether this measure should be given another name. We chose to label this as “compliance” to be consistent with previously published papers on the issue, but the reviewer is correct in pointing out that we do not differentiate between compliance and preventive behavior. As we see it, the engagement in preventive behavior (“...washing hands, avoiding travel and situations with other people, keeping distance, and avoiding touching things”) and complying with the infection control measures overlap considerably. We believe that “preventive health behaviour” may be a more general term than what we mean by intending to comply with the specific COVID-19 infection control measures that were recommended in Norway at the time of data collection. Since “compliance” in our study is measured by an item asking about people’s intention to follow the infection control measures suggested by the authorities (we then listed these), we think it is best to label it as compliance and not “preventive behaviors” or “preventive behavioral intentions”. We have tried to make this distinction clearer in the manuscript by changing the following to section 1.1.2 Compliance:

- *We have changed the headline in the introduction of compliance from “Compliance” to “Compliance to infection control measures” as it better captures the specific type of behavior we intended to study.*
- *We have taken the discussion of preventive behaviors and compliance into account and clarified the relationship in the introduction. The first part of section 1.1.2 now reads as follows: “Compliance with infection control measures is a key component of preventing the spread of infectious diseases. Measures such as hand hygiene, social distancing, and use of protective equipment, are essential for protecting society at large as well as individuals who are at particular risk. However, it may be challenging for individuals to comply with numerous measures to mitigate infection spread as it often requires individuals to change their behavior in inconvenient ways and adopt new habits. In the context of COVID-19,*

compliance with infection control measures and engagement in preventive behavior are intricately intertwined, often overlapping considerably. Both concepts entail actions aimed at reducing the spread of the virus and minimizing individual and collective risk. Compliance with measures such as wearing masks, practicing hand hygiene, and maintaining social distancing constitutes a proactive approach to preventing transmission, aligning closely with behaviors typically associated with preventive action. Consequently, in much of the research on COVID-19, these terms are frequently used interchangeably to represent the collective efforts individuals undertake to mitigate the impact of the pandemic (Brouard et al., 2020; Burton et al., 2021; Clark et al., 2020; Harper et al., 2021)."

- *We have added information about why we may expect a dynamic relationship between compliance to infection control measures and perceived risk (see last paragraph under section 1.1.2.*

Fifth issue: Participants and power analysis

Reviewer 1 stated "Third, It is not clear which participants would be retained (only participants who responded to all four waves?) and no power analysis is carried out to ensure the relevance of the sample available." The last part, again, could be fixed with the providing of a syntax/analytical code and lowering the threshold for significance with a rationale associated. However, authors also need to be more specific around the exact sample used especially because the sample size for T4 is more or less half of T1.

Response:

We agree that the analysis plan suggested in phase 1 of the manuscript was unclear about how many participants that would be reattained for the final analysis. We have taken steps to improve the clarification of this:

- *We have decided to suggest a new model that only uses the participants that have responded to all rounds, keeping only the panel sample for testing the hypotheses. With the attrition rate we expect from this panel, this would constitute $N \sim 2000$.*
- *The reason why T2 and T4 are dramatically smaller than T1 and T3 is not due to attrition, but due to different collection plans. In T2 and T4 the survey sample was divided between several research groups, that each asked their questions. Each research group is given approximately a panel sample of 2000 participants for their questions that they can observe over time. However, in T1 and T3, a larger share of the participants received questions related to COVID-19. We have tried to make this clearer in the manuscript under sections 2.1 and 2.2. Nevertheless, all samples should be representative and large enough to reliably test the hypotheses. We have tried to make it clear in the description of the participants:*

"Data for this study were collected in four rounds, in March (T1, $n = 4083$), June (T2, $n = 2820$), August-September (T3, $n = 5541$) and November (T4, $n = 2533$) in 2020. We assume that most (90%) of the participants remain from one round to the next based on prior tendencies in the Norwegian Citizen Panel data. The core panel sample size is around $N \sim 2000$. However, for rounds T1 and T3, the survey was fielded to a larger share of the Norwegian Citizen Panel, resulting in a sample size of 4083 and 5541 for these rounds respectively."

- *As the editor mentions, the uncertainties will be partially solved by providing the analysis script and lowering the threshold for significance as described under issue three above.*

Reviewer #1:

The research question is not really new, but it is interesting and may be of some importance in establishing potential bidirectionality between risk perception and compliance with preventive behaviour on a longitudinal and (fairly) representative database.

The hypotheses are credible and precise. However, hypothesis H3 should be introduced and justified.

The data has already been collected and the targeted measures from the first wave have been used in three manuscripts (published or pre-printed). In one of these manuscripts, the authors even test the link between perceived risk and compliance. Although the authors state this, it is a concern to me. Indeed, the authors have already tested H1a, so the level of prior knowledge of the data is not satisfactory. The authors state that it is not the same items that have been used; but this poses a limit to the measurement if the authors consider that it is not the same concepts, or if it is conceptually the same thing, then the authors have already tested H1a.

Second, there are discrepancies between the theoretical reasoning and the methodology used. In particular, when perceived risk is defined, it is defined as a risk to oneself. On the other hand, it is measured using several items relating to the risk associated with the disease for one's own health, that of others, and even for one's life in general. This should be specified if and why differences in effects can be expected. Second, and similarly, compliance involves a range of different behaviors. Some authors have argued for a distinction between different preventive behaviors (e.g., distancing from hygiene). The authors should have taken this into account in order to shed more light on their research question.

Third, It is not clear which participants would be retained (only participants who responded to all four waves?) and no power analysis is carried out to ensure the relevance of the sample available. To conclude, I have serious reservations about whether these hypotheses, with these data, can be tested in an RR format. This does not preclude the interest of the question, but rather the appropriateness of the publication process.

1. First issue: H3 justification

Hypothesis H3 should be introduced and justified.

Response:

Thank you for bringing the issue of H3 to our attention and we have handled it as described to the editor in "First Issue: H1a and rationale for the hypotheses" above.

We agree that it should be clearer in the manuscript why we expect a relationship between compliance and perceived risk, and why it is unclear in what direction it goes throughout a year. We agree that H3 would be stronger if we predicted the direction of the relationship. We chose not to do this as perceived risk and compliance are rarely studied longitudinally with a panel, and the reverse relationship, with compliance as the predictor and risk as the outcome, is even less documented in the literature. Investigating this relationship without assuming a direction of the relationship may better represent the current "state of knowledge", as compliance to preventive measures and perceptions of risk may influence each other dynamically over time. We concede that it is more plausible to think that higher past compliance may be related to a reduction in perceived risk, but we do not feel sufficiently confident in the evidence to justify this direction. We are, however, open to changing this if the reviewers think this would improve the study.

2. Second issue: H1a

The data has already been collected and the targeted measures from the first wave have been used in three manuscripts (published or pre-printed). In one of these manuscripts, the authors even test the link between perceived risk and compliance. Although the authors state this, it is a concern to me. Indeed, the authors have already tested H1a, so the level of prior knowledge of the data is not satisfactory. The authors state that it is not the same items that have been used; but this poses a limit to the measurement if the authors consider that it is not the same concepts, or if it is conceptually the same thing, then the authors have already tested H1a.

Response:

Thank you for bringing this to our attention. We completely agree that this was a significant issue and we have handled it as described to the editor in “First Issue: H1a and rationale for the hypotheses” above.

3. Third issue: Risk and compliance measure justification

There are discrepancies between the theoretical reasoning and the methodology used. In particular, when perceived risk is defined, it is defined as a risk to oneself. On the other hand, it is measured using several items relating to the risk associated with the disease for one's own health, that of others, and even for one's life in general. This should be specified if and why differences in effects can be expected. Second, and similarly, compliance involves a range of different behaviours. Some authors have argued for a distinction between different preventive behaviours (e.g., distancing from hygiene). The authors should have taken this into account in order to shed more light on their research question.

Response:

Thank you for bringing this to our attention. We agree that the theoretical account of the concept and its operationalization should be aligned. We have handled this issue as described to the editor in “Fourth issue: Definition and operationalization” above.

Note also that we have revised the analytical approach to include a multiverse analysis. This will partly resolve the issue of operationalization by testing the model with all possible operationalizations of the “Perceived risk” variable. We will thus see how the risk judgments for self, others, and situational changes may impact compliance with infection control measures.

4. Fourth issue: Participants

It is not clear which participants would be retained (only participants who responded to all four waves?) and no power analysis is carried out to ensure the relevance of the sample available.

Response:

Thank you for bringing the issue to our attention. We agree that this was not handled in a clear way, and we have made adjustments to the analysis plan as described in our response to the editor under “Fifth issue: Participants and Power analysis” above.

Reviewer #2:

Thank you for the opportunity to review the Stage 1 report of “Relationship between perceived risk and compliance to infection control measures during the first year of a pandemic”. The registered report proposes to test the association between risk perception and compliance with covid-19 control measures among a representative Norwegian sample, while also testing how this association changes over time. The report is concise, well-structured, easy to read, and clearly spells out the contribution to the literature. This is my first time reviewing a registered report based on secondary data. The authors are very transparent in reporting prior knowledge with the data that they plan to use. Overall, I think this report generally meets the Stage 1 criteria, but I have some minor comments and suggestions that I hope will be useful.

I wish the authors good luck with their paper!

1. First issue: Statistical threshold

The authors specify using a p -value cut-off of 5%. But given the very large sample size (in each time point), I wonder whether this cut-off should be reduced. Some have argued that when sample size is very large, p -values right under .05 (e.g., .04) might be taken as evidence for the null (e.g., see: <https://journals.sagepub.com/doi/10.1177/25152459221080396>).

Response:

Thank you for bringing the issue to our attention. We have reduced the threshold for significance, and we have made several other adjustments to the analysis plan as described in our response to the editor under “Second issue: Risk of bias” above.

2. Second issue: Statistical approach

This brings me to my next comment: In Table 3, if I have interpreted this correctly, the authors specify that lack of support for a hypothesis will be interpreted as indicating the absence of an association. This means that p -values right below 5% would be taken as evidence for an association when in fact they might indicate evidence for the absence of an association. I’ve never worked with sample sizes this large, and I have no experience with the methodology proposed in this study, but I wonder if one might want to complement traditional null hypothesis significance testing with alternative methods that quantify evidence for the null more directly (e.g., equivalence testing or Bayesian analysis).

Response:

This is an interesting and important point, and it is particularly worth discussing in the interpretation of the results. We’d like to specify that according to the NHST framework, we will not consider the absence of proof to be proof of absence. As previously noted, we have changed the inference criteria and will now perform a multiverse analysis on the relationships H1-H3. Together this will give us a stronger basis for making inferences about the strength and magnitude of the relationships (or lack thereof) in this study. We do not think that alternative methods would add to this at this stage.

3. Third issue: Directionality

On page 6, the authors write that they will also test the reversed relationship, that is, whether compliance predicts perceived risk. Although I think I know why the authors want to test this, I think it might be good to provide a brief justification, and perhaps to also explain why they want to test it non-directionally.

Response:

Thank you for bringing the issue of directionality to our attention. We agree that H3 would be stronger if we predicted the direction of the relationship. We have described our response to the editor under “First issue: H1a and rationale for the hypotheses” above.

Fourth issue: Results section with simulated data

Just a suggestion and I don’t mean to complicate things, but it would be helpful to see the whole Results section written out based on simulated data and share the analytical code (by uploading it to the OSF page). This would make it easier to compare the Stage 1 report with the Stage 2 report and would allow the authors to potentially get feedback on their analysis script and detect potential errors.

Response:

As far as we understand, the PCI RR stage 1 submission format is not supposed to include a Results section with placeholders for analysis results, or to show analyses based on simulated data. We have included an analysis script that has been tested on simulated data, see response to the Editor’s third issue.

4. Minor issue

This is a very minor suggestion, but I think this part on page 6 can be removed: “We will use a registered report publication process to enhance the transparency and rigor of our research methodology, study design and analysis plan. This approach ensures that the significance of our study is evaluated based on the research question and methodology, rather than the outcomes.”

On page 6 there is a missing closing parenthesis (in “(H3”).

Response:

Thank you for bringing this to our attention. We agree that it was unnecessary to include and in response we have shortened the section to briefly spell out our approach in the current study (See 1.3 Current study): “The current study used a registered report approach (Grand et al., 2018).”

Reviewer #3:

The authors propose a study of the relationship between risk perception and compliance with infection control measures during the first year of the COVID-19 pandemic. The study uses a large representative sample of the Norwegian population, with measurements at four time points, between March and November 2020. This represents a great opportunity to investigate how the relationship between risk and compliance develops over time. I find the proposal timely, interesting, and generally well planned, but I have some points which I think could be improved, mostly concerning methodology and analysis.

I find the research question to be interesting and scientifically justifiable. The introduction is well-written and gives a clear rationale for the proposed study, and the study connects well to previous research on similar topics.

There are three hypotheses. H1 proposes a positive association between risk and compliance at each time point. H2 proposes risk at previous time point will be positively associated with compliance at the subsequent time point. H3 proposes that compliance at previous time point will predict risk at subsequent time point. All three hypotheses seem logical and plausible.

1. First Issue: Discuss rationale for hypotheses in introduction.

However, even though the introduction gives a nice overview of the topic and the need for the proposed study, I feel like there is room to expand a little bit concerning the hypotheses and the potential outcomes of the analyses. For instance, in the study design template in Table 3, the authors discuss some interesting interpretations of different outcomes. Especially for hypothesis 2 and 3, I think these could be discussed in the introduction. Communicating for example the idea that high risk could lead to establishing habits that persist over time and/or could lead to fatigue would help readers think about the importance of the research and about potential implications, beyond what is already included.

Response:

We agree that the hypotheses could have been more clearly argued for in the Introduction, and that the ideas of establishing habits and being fatigued should be made clearer. In response, we have changed the last paragraph under section 1.1.2 about compliance to infection control measures to include a discussion of why the relationship between “compliance” and “perceived risk” may change during a pandemic. The paragraph now reads:

“As the pandemic progressed, perceptions of risks related to COVID-19 may have changed for some individuals. Some may have become complacent about the situation, while others may have become more fearful as the number of cases rose and new variants of the virus emerged. Through the first year of the pandemic, new information and guidelines were periodically released, and this may have affected people’s confidence in and motivation for complying with the infection control measures. On the other hand, people may have gained confidence in the control measures as it may have prevented them from getting infected with COVID-19, and they may have ascribed this to the effectiveness of precautionary behavior. As a consequence, compliance with infection

control measures may contribute to perceiving less risk of getting infected with COVID-19.”

We also made an addition to section 1.2 “Knowledge gap” to reflect why we are uncertain about the direction of the relationship between prior compliance and perceived risk. The addition reads as follows:

“The combination of perceived risk and compliance to protective measure are rarely studied longitudinally, and the reverse relationship, with compliance as the predictor and risk as the outcome, is even less documented in the literature. While we should expect perceived risk to predict compliance, less is known about how past engagement with protective behavior may impact how people perceive the risk scenario.”

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

As acknowledged by the authors, the fact that the data has already been collected and that the authors have already inspected some key measures and tested the association between risk and compliance at T1, makes the proposal a “Level 1” submission. Even though the key research questions have not yet been analyzed, I believe the authors should think more about how they can further reduce the risk of bias.

2. Second issue: H1a

The authors explain that they have previously described the association between risk and compliance at T1 in a couple of publications, so that they have in a sense already tested H1a. Even though the association they tested “also included several other items not included in the present analysis”, it seems important to (briefly) describe what this previous analysis showed. From a quick look at the publications, it seems like the association between perceived risk and compliance was not straightforward at T1. This should be described so that readers are made aware of what the authors already know about the relationships that will be investigated in the current proposal.

Response:

See our response to editor’s first issue above. We have added the following to section 2.3:

Firstly, response distribution for perceived risk and compliance at T1 has been examined and reported (Sætrevik, 2021) and we have previously described the cross sectional association between perceived risk and compliance at T1 (Sætrevik & Bjørkheim, 2022). Contrary to our hypothesis, we found that those who perceive a higher risk are actually less likely to comply with safety measures. When we specifically looked at the risk of getting infected, we found a very small negative effect on compliance. However, when we asked about the perceived risk of infection for the general population, the effect on compliance was small and positive. Overall, our findings only weakly support the idea that seeing the pandemic as a threat leads to following safety measures, and only when considering the risk to the general population, not to oneself (Sætrevik & Bjørkheim, 2022). This association will not be considered among confirmatory hypotheses in the current study.

3. Third issue: Reduce bias for Level 1

submissions is to have “an extremely conservative statistical threshold”, but the authors suggest in Table 3 to use the conventional .05 level. I think it would be good to consider a stricter statistical threshold. Given the large sample size, the authors would probably have power to detect relatively small effects even if the alpha level is set at a stricter level. If the authors decide against this, they should at least explain why they believe the conventional threshold is sufficiently strict enough.

Power Analysis: On a related note, the authors do not report a power analysis. It would be interesting to know something about which effect sizes one could reliably detect with e.g. 90 or 95% power. Although I suspect this is complex for the proposed model, it seems that there are packages that could be helpful: <https://www.tandfonline.com/doi/full/10.1080/10705511.2022.2122467>

If it turns out that it is too complex to perform meaningful power (sensitivity) analysis, it would still be helpful to include a section discussing the topic in relation to the sample size and potential effect sizes.

Third, another suggestion to reduce bias is to recruit a blinded analyst. I am not sure if this is necessary here, but I include the point to illustrate that the authors should emphasize the issue of potential bias to a greater degree. Expanding on section 2.3 and explaining which countermeasures they have taken, or why such countermeasures are not necessary, would be helpful.

Response:

Thank you for bringing this to our attention, and we agree with the need for more stringent countermeasures against bias (lower p-value, multiverse analysis and blinded analyst). See our response to editor's second issue above.

Concerning the comment about performing a power analysis:

- The paper suggested by reviewer 3 gives a good introduction of how one can perform a power analysis based on simulated data. However, to perform such an analysis we would have to make a number of assumptions about the between and within unit variation in order to provide a meaningful power estimate. We do not have sufficient knowledge of between-unit variation to confidently make those assumptions. 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 $n \sim 2000$ in our study, and we should thus expect to have a power of more than .80 to detect "small" cross lagged effects.*
- Note also that the data collection is completed and a post-hoc power analysis cannot inform our decisions on sample size at this point.*

4. Fourth issue: Analysis pipeline

The analysis plan is clearly, but quite briefly described. In light of the issues discussed above, I think it would be good to provide syntax for the proposed analysis. This would help to minimize concerns about analytical flexibility.

I also have a question. I do not know much about the kind of model proposed here, but regarding for instance H2, I wondered whether the proposed positive association between e.g. T1 risk and T2 compliance controls for perceived risk at T2? This might be a terribly naïve question, but I think some more details about the analysis plan could help those who are not well versed in cross-lagged panel models to get a better picture of what is going on.

Response:

See our response to editor's third issue above, where we mention the analysis script. As for the second issue raised here, we've also had some similar discussions in our group. Our understanding is that the various associations tested within a cross-lagged panel model will control for each other. Thus, for each association the model will report the "remaining variation" that is not accounted for by the other associations in the model.

5. Fifth issue: Data and code

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.

As mentioned, code/syntax for the proposed analysis would be beneficial. Additionally, it is not clear whether it will be possible for others to access the data. It would be best if the data could be shared openly, but I suspect there might be privacy concerns here preventing this possibility. If so, it is important to state this clearly, and to discuss measures that would ensure maximum transparency and reproducibility.

Response:

An anonymized version of the dataset (without the personal identification code) will be provided in full and shared on the OSF page (<https://osf.io/5k7qw/>) of this project. As mentioned above we will also provide the statistical script such that everyone can replicate the analysis. Some countermeasures to enhance privacy are inherent in the data. For instance, all background demographic variables are measured on a group level in our data. For example, age is measured as “belonging to an age group within a ten-year span” and not exact age). We have added a more detailed description to the manuscript about what kind of data are shared, in the following text in section 2.1 Participants:

“An anonymized version of the dataset, devoid of personal identification codes, will be made available and disseminated via the project's Open Science Framework (OSF) page (<https://osf.io/5k7qw/>). Various precautions aimed at bolstering privacy are inherent in the dataset as all background demographic attributes are measured at a group level.”

Other comments

I found Figure 1 illuminating. The fluctuations in the infection rate also reveals something about the “objective” risk at different times in the pandemic (with some important caveats, e.g., the high uncertainty especially at T1 may make it difficult to give an objective risk measure). I wonder whether it would be interesting to control for “objective” risk levels, by using infection rates or something similar. Or perhaps one could compare effects of perceived risk on compliance with effects of objective risk? I'm not sure if this is at all feasible, but raise this question as something the authors can consider.

Response:

Thank you for your thoughtful suggestion. We agree that controlling for objective risk levels could provide valuable insights into the dynamics of perceived risk and compliance, particularly in understanding how individuals respond to varying levels of risk perception compared to actual infection rates. However, incorporating objective risk measures into our analysis may introduce complexities and potentially confounding variables that could complicate the interpretation of hypotheses. While we recognize the theoretical importance of considering objective risk levels, we believe that attempting to integrate these measures into our model at this stage may result in methodological challenges that outweigh the potential benefits.

As shown in Figure 2, we have no hypotheses about how risk levels will change between T1, T2, T3 and T4 (which would perhaps be most influenced by changes in infection rates). Instead, our hypotheses test the extent to which that risk accounts for compliance at the same time (H1), compliance in the future (H2) and the relative amount of that change that can be accounted for by compliance in the past (H3). We believe this would be less subject to how e.g. different infection rates

over time influence how risk is seen at a given time. Nevertheless, by presenting the average level of risk at each time point, will also make clear how that changes over time, and we will discuss how that may be influenced by factors such as objective infection rates, and explore possible avenues for addressing this issue in future research. We agree that this is an interesting avenue for further investigation but at this stage we are not confident that such a considerable change in design will improve the study sufficiently.

Minor comments:

Abstract: “during the first year COVID-19 pandemic in Norway” à “during the first year of the COVID-19 pandemic in Norway”

p. 3 “They also found that psychological factors to be” à “They also found psychological factors to be”

p. 5 “The current study may help to fill the knowledge gap” à I think you can allow yourselves to be less modest here.

Response:

Thank you for pointing this out. We have edited the text as suggested.

Conclusion

I think the topic is very interesting and worthwhile to pursue, and that the current proposal is solid. My main comment is that I see potential for improvements when it comes to controlling for potential bias. I hope the authors find my comments helpful.