Personality and compliance – response to first PCI RR peer review

Peer review received date: 19.03.2024

Manuscript revised date: "encouraged to revise promptly"

Editor: Andrew Jones

We would like to thank the editor and the two reviewers for taking the time to review our Stage 1 manuscript, and for all their relevant and important comments. We are sorry that it took us longer to return a revised version of the manuscript than we would have liked.

We have revised the manuscript based on this input, and we feel 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, and the action we have taken based on each issue.

In addition to the issues raised by the reviewers, we have also made a number of minor adjustments to the text throughout the manuscript to increase understanding, flow, and readability. Note that we have adjusted author order on the title page to be aligned with the author order listed that was already listed in the PCI submission portal (order BS, EKE, SBB). All authors are aware of and in agreement about this update.

Respectfully on behalf of all three authors,

Bjørn Sætrevik

Editor

1. I would report a different measure of internal consistency rather than alpha (such as McDonald's omega - https://journals.sagepub.com/doi/full/10.1177/2515245920951747).

Response:

Thank you for highlighting the concern regarding internal consistency. We agree that McDonald's omega offers a more precise estimate of internal consistency, especially in situations with few items, compared to Cronbach's alpha. We will report McDonald's omega instead of Cronbach's alpha to assess the internal consistency in the personality variables. This is now stated in the third paragraph of section 2.4: Analyses:

> *"We will use McDonald's omega for internal consistency of the items."*

2. How might you consider any issues with multicolliniearity in your regression models?

Response:

Thank you for pointing out that we should describe how we will consider issues with multicollinearity. In order to do so, we have now added the following text to the methods section:

> *"Bivariate correlations will be reported and investigated in order to describe the data and to test for multicollinearity. Correlations above .7 between independent variables will be interpreted as multicollinearity as recommended in Pallant (2020). In addition, the tolerance and variance inflation factors will also be investigated to check for multicollinearity in which tolerance factors*

below .10 or variance inflation factors above 10 will be interpreted to suggest issues with multicollinearity (Pallant, 2020)".

Reviewer #1

1. First and foremost, it is unclear whether original data from either datasets were already analyzed for different purposes and/or whether they are currently not accessible to the authors - I find this rather unlikely consider the large number of respondents. As the authors mention that the two datasets were not combined yet, how will they evaluate whether the final sample size will be sufficient to test their predictions?

Response:

It may not have been clear in the previous version of the manuscript what we already knew about the the two extant datasets that compose the datasets for the planned study. All the data was collected as part of a nationally representative panel survey that has been running for several years. The panel data is collected as a joint collaboration between various teams of researchers associated with the University of Bergen. Different researchers (that are only vaguely aware of each other) work on the panel data at different times, they suggest and analyze different parts of each data collection round (sometimes split up into sub-samples), based on various research questions. The data from the various data collection rounds are stored on a central server, and individual researchers get permission to export portions of the collected data to analyse the research question of their interest.

The personality dataset was designed and collected before the current authors were involved in the panel data collection. As far as we know, this dataset was not originally intended to be combined with health behaviour variables. We have downloaded the personality dataset from the central server, but we have not examined any of the variables from that dataset. We have extracted a selection of information about the dataset (such as sample size) in order to provide a minimal description in the Stage 1 manuscript. To our knowledge, the personality data have not been analysed yet, and we have not identified or been in contact with the researchers who collected these data (who may have done some preliminary analysis).

Two of the current authors (BS and SBB) designed the pandemic survey questions for the August 2020 panel data collection (repeating some items first included in March 2020). These pandemic variables have been explored and reported from that dataset to some extent. We mentioned this in the previous version of the manuscript, but have now tried to make it clearer, in statements such as this (section 1.2.1):

> *"a previous analysis of other data from the current panel (measured in March 2020) failed to show a substantial association between perceived risk and compliance (Sætrevik & Bjørkheim, 2022)."*

And in section 1.2.2:

> *"Our previous analyses of compliance in the current dataset have shown that compliance was very high among Norwegians in March 2020 (Sætrevik, 2021), decreased somewhat over the summer as infection rates fell, but rose in the early fall when the rates increased again (Bjørkheim & Sætrevik, 2020; Sætrevik & Bjørkheim, 2020)"*

Thus, both datasets that we will use in the current paper are already available for us. We have some knowledge about the pandemic dataset, but very limited knowledge about the personality dataset.

Crucially, the two datasets had never been aligned, and no comparison has been made between the personality data and the pandemic data. We think that it is a strength that we have not combined the datasets, as this prevents us from doing preliminary analyses which might influence our hypotheses and analytic strategies. Our rationale for not combining the datasets are now described in the report with the following text at the end of section 2.1:

> *"The dataset with personality measures and the dataset with pandemic measures were not merged before Stage 2, in order to prevent preliminary analyses from affecting the hypotheses".*

As for the reviewer's question about what combined sample size can be expected and whether this will be of a sufficient size, we have made an estimate based on the attrition rate between other rounds of the current panel dataset. In the previous version of the manuscript, we assumed that the sample of the current study, consisting of those who participated in both relevant data collections would amount to n = 3,000. We have now adjusted this to 2,000, in order to be very conservative. This is nevertheless sufficient to observe what we have set as the "smallest effect size of interest". The G*Power calculation to show the sensitivity of the analyses has been placed in the OSF folder for the project (https://osf.io/g57sf/), and is described in Table 2 of the manuscript. We have also added the following text to section 2.1.2:

> *"Of these, [number to be added at Stage 2] panel members could be matched between the personality and the pandemic data collection [exact number not known at Stage 1 as the datasets have not yet been combined. Given typical attrition rates in this panel, we assume the combined dataset will be n > 2,000, which should be sufficient for the planned analyses (see Table 2 for sensitivity analysis)]."*

2. In general, I find all planned analyses and predictions rather exploratory - maybe this is confounded by the large number of hypotheses. The design table does not help much with this since it is quite crowded and could probably be simplified.

Response:

We were somewhat confused by this comment, which may come down to different ways of using the term "exploratory". In the usage we are familiar with, "exploratory research" is data-driven and hypothesis-generating, as opposed to "confirmatory research", which is pre-planned and hypothesis-testing. We would argye that the current research is the latter, using a registered report format to announce the hypotheses to be tested in advance, in order to separate that from anly exploratory analyses that may be added later.

It could be that the reviewer means that the research is "exploratory" in the sense of planning to test (uni-directional) hypotheses for most of the possible associations between traits and pandemic outcomes, and that some of these associations may have a weaker empirical basis than others.

This approach was chosen in order to be able to perform confirmatory tests of all the relevant associations that we found to be reasonably expressed in the literature. The addition of Table 1 may make the hypotheses more easily accessible (see below). In section 3.3-3.10 we have expressed what support each of the hypotheses has, and which of these we feel can be stated with more or less confidence. While we acknowledge that this may make the manuscript somewhat long and cumbersome, we also feel that it would also be a shame to forego the opportunity to test some of the relationships. If we were to remove some of the hypotheses, we could end up with results that

support patterns that are aligned with theory and empirical findings from the literature we cite, that we would nevertheless have to interpret as "novel findings". We have now added explicit statements when introducing the hypotheses in section 1.5 about the support each of them has in previous literature:

> *"Based on a number of previous studies, we expect (H1) Conscientiousness to have a positive association with Compliance. There is also considerable support for an expectation that (H2) Agreeableness will have a positive association with Compliance. We expect (H3a) Extraversion to have an inverse association with Perceived risk, although there is limited support for this in previous literature. There is more literature to support the expectation that (H3b) Extraversion will have an inverse association with Compliance (although previous literature may only support this for measures related to social interaction). Although previous literature on this is conflicted, we lean towards expecting (H4a) Openness to have a positive association with Perceived risk. Previous pandemic literature provides some reason to expect that (H4b) Openness will have a positive association with Compliance. Based on general descriptions of the trait, we may expect (H5a) Neuroticism to have a positive association with Perceived risk. Finally, based on a few recent studies we expect (H5b) Neuroticism to have a positive association with Compliance."*

We think it would be difficult to argue that studies like the current one has "too many hypotheses". The literature we cite use null-hypothesis significant testing approach (with a notable exception of Han, 2021), and they measure and presumably test relationships between all big-5 factors and typically at least two outcome measures. Some of these studies do not explicitly state their hypotheses, or state summary hypotheses that include several of the outcome measures. Thus, we should assume that these studies have used at least as many hypotheses as we suggest in our study, although they have not been explicitly stated, or only the ones that ended up having positive results are stated. We prefer to avoid steps like collapsing two outcome measures into one hypothesis (e.g., turning H3a and H3b into a joint H3), because we could then end up with results showing both support and non-support for a given hypothesis.

Note also that the second reviewer makes the opposite point, that we should also include predictions for all personality traits to both possible outcome measures (i.e., also the two associations not covered by the current hypotheses).

We can agree with the reviewer that the design table (Table 2) is somewhat overwhelming and can be difficult to manuver. The inclusion and formatting of this table is "strongly recommended" in the PCI RR format, and we do not think it can be adjusted too much. However, we have added a simpler Table 1 to the manuscript (see copy below) that may make the hypotheses clearer.

	Personality factor	Direction	Outcome	Support in literature
H1	Conscientiousness	Positive	Compliance	Strong
H2	Agreeableness	Positive	Compliance	Strong
H3a	Extraversion	Inverse	Perceived risk	Limited
H3b	Extraversion	Inverse	Compliance	Moderate
H4a	Openness	Inverse	Perceived risk	Limited
H4b	Openness	Positive	Compliance	Moderate
H5a	Neuroticism	Positive	Perceived risk	Limited
H5b	Neuroticism	Positive	Compliance	Moderate

3. The authors state in the Design Table (but it seems not in the main text) *"Due to a large sample size, even small effects are likely to show significant effects. Due to small effects having impact on infections in large populations, we will consider effects larger than f2 = 0.01 to be theoretically meaningful for our research question."* How will the authors rule out that is not just another "statistical artefact" similar to the ones mentioned in the introduction? What do they mean by "theoretically meaningful" when their earlier consideration is in fact very practical (i.e., small effects having impact on infections in large populations).

Response:

We agree that the use of the term "theoretically meaningful" was unprecise. We have now changed this to "meaningful from a public health perspective".

We see that our attempt to raise two different issues in this section was confusing. The first issue was to argue that small effects could be interesting in this setting, while the second issue was the need to set a "smallest effect of interest". We also see that such discussions could be removed from the table and instead placed in the paragraph text. There was already a similar statement in the second paragraph of section 2.4: "Analyses", which has now been changed and expanded:

> *"Nevertheless, under a pandemic with exponential infection rates, small effects that changes the behaviour of a few people can have a large impact on the pandemic's development in the population. This may protect many people from infection, and impact health outcomes for people at-risk. On the other hand, we should also be aware that arbitrary variation may produce significant effects in large sample sizes. We will therefore set a 'smallest effect size of interest' at $f^2 = .001$."*

After removing the overlapping segments, the relevant part of Table 2 now reads:

> *"We will consider effects larger than f2 = 0.01 to be meaningful from a public health perspective."*

The reviewer also raises the issue of ruling out statistical artifacts in our study. While such issues may be difficult to completely rule out, we have taken additional steps to reduce the risk of Type 1 error. Motivated by the reviewer's comment, we have lowered the alpha level from $p < .05$ in the previous version, to $p < .01$, and made a more conservative estimate of the sample size (from n = 3,000 th n = 2,000).

4. In general, it does not seem that the aim of the report and the applied methods derive clearly from the introduction - which at the moment is a rather disconnected list of previous research on various factors that might or might not correlate with each other. The report should make transparent to the reader how their predictiond and planned tests are answering specific questions and gaps in the literature.

Response:

Our intention with the Introduction was to first introduce the two pandemic outcome variables and all five personality factors, then review the previous literature on how they have been shown to be associated, and finally use that to state our hypotheses. Based on the reviewer's comment (and a similar comment from the other reviewer), we realize that this structure may have been overly complex, and we have now dramatically reorganized the Introduction. We now present each

personality factor in turn, and present the previous literature associated with theat factor at the same time (sections 1.3.3 to 1.3.10). We hope this structure will be easier to follow for readers.

The section "Knowledge gap" has also been comprehensively revised to address the issue raised by the reviewer. This has been done to show how we think the past literature can be improved on, and how our study can contribute to this. Some of this consists of content moved from sections 1.3.3 to 1.3.10, and we have added two sentences to section 1.3.2 to anticipate this:

> *"All the big-5 personality traits have been indicated to be involved, but the indications are clearer for some traits than for others, and some of the associations have little or contradictory support. There are more studies about the association personality traits have to compliance than about the association they have to pandemic risk perceptions."*

This has resulted in a substantially expanded section 1.4, "Knowledge gap", from one to five paragraphs. Here we discuss issues related to variability in the support for associations between personality traits and pandemic outcomes, that cross-sectional studies where both personality traits and pandemic outcomes are measured at the same time may be subject to normative influences, to mood effects, and order effects, and that findings from non-registered studies may be less robust.

We have also added more background in order to make it clearer how the different elements that are introduced are expected to connect. Among other such changes, we have added the following text to section 1.2.1:

> *"How people see risks during a pandemic may be shaped by various psychological mechanisms. For instance, iIndividuals may rely on their past experiences with infectious diseases, information from various news sources, and their prior beliefs to evaluate risk. In addition, factors such as personality traits, level of trust in authorities, and cultural beliefs may influence how people perceive and respond to risk. For example, someone with a higher tolerance for risk might perceive the threat of the virus differently than someone who is more risk averse. Similarly, individuals who trust government guidance advice may be more likely to trust infection rates and take precautions compared to those who are sceptical of government information (Ebrahimi et al., 2021; van der Weerd et al., 2011)."*

5. Finally, I find some statements rather speculative *"The results may facilitate how future pandemics are handled, in particular in terms of adjusting public health information to be effective for reaching individuals with personalities that may otherwise be resistant to seeing the risk or to comply with infection control measures"* and should be significantly toned down. Similarly this statement *"Perhaps due to a sense of urgency, most of the research on how personality may influence pandemic behaviour was not performed in accordance with current standards for open and transparent research, (i.e., controlling the degrees of freedom in measurement, analysis, and hypothesis development). If measuring a number of personality traits along with several different pandemic attitudes, beliefs or behaviours (which may be indexed in different ways) in large cross-sectional studies, a high number of potential relationships can be discovered. This makes it difficult to discriminate true psychological mechanisms from spurious false positives findings that may emerge*

from multiple comparisons and undisclosed analytic flexibility" would let the reader think that new data are going to be collected, but then they found out this is not the case. Rather, what they later gather from the paper is that the datataset involved in this RR was indeed collected with the same procedure it criticizes.

Response:

We have changed the first text quoted by the reviewer to be more concrete about what the study may contribute (in the first paragraph of section 1.1):

> *"Knowledge about how personality traits and other individual differences determines risk perception and compliance may be relevant for designing public health interventions. In particular, information campaigns may be adjusted in attempts to influence individuals that may otherwise be resistant to seeing the risk or complying with infection control measures."*

In the second text quoted by the reviewer, the intention was now to give the impression that new data would be collected after the Stage 1 in-principle-approval. Our intention was to raise the issue that without clear distinction between the hypothesis-generating and the hypothesis-testing phases of the research process, it may be difficult to be confident that the formulation of the hypotheses is not influenced by the analysis process. This issue is the reason why we have chosen a "registered report" approach for the current study. While the reviewer is right that all the data for the study has already been collected at the time of the stage 1 manuscript development, it has not yet been analysed (i.e., compared between the two time points of data collection, see more about this in our response to the first issue raised by the reviewer). Various checks in the registered report format are intended to control for the potential bias of the data already having been collected. We would therefore argue that our data was not collected "with the same procedure we criticize", since we are taking quite explicit steps to separate exploration from confirmation. We have now tried to make our point clearer in the fourth paragraph of section 1.4 (and changed from a negative to a positive framing). This paragraph now reads:

> *"Relatively few of the studies on how personality may influence pandemic behaviour have separate research procedures to distinguish hypothesis statement from hypothesis testing. Although this may be understandable for research initiated during an ongoing health crisis, this may make it difficult to say how robust the findings are and what predictive value they have. When measuring a number of personality traits along with several different pandemic outcomes (attitudes, beliefs, or behaviours, which may be indexed in different ways) there is a high number of potential relationships that can potentially be discovered. An approach where the planned hypotheses are registered in advance of the analysis can make stronger claims about whether a priori predictions are supported (as opposed to false positives findings that may emerge from multiple comparisons and undisclosed analytic flexibility, Munafò et al., 2017; Nelson et al., 2018)."*

Based on this comment, we have also edited and moved one of the sentences in the Abstract:

> *"To provide transparency and to control for flexibility in the analysis and reporting of the many possible associations, hypotheses and analysis plans were reviewed and approved in advance of aligning the two datasets (a registered report format)."*

Reviewer #2

1. The literature review in the introduction could be expanded on more. The authors should review additional work that examined personality and individual differences and pandemic responses during the Covid-19 pandemic. This can then help pinpoint why the additional examination of the Big Five trait would be relevant, especially in the face of other research that as explored different psychological mechanisms. Basically, why do personality differences matter? What can the current work contribute to the existing literature? The journal Social and Personality Psychology Compass also had a few special issues on social/personality psych and the pandemic, and it would be good for the authors to look at some of the published findings there to include in their literature review, e.g., Panish et al. (2023; https://compass.onlinelibrary.wiley.com/doi/full/10.1111/spc3.12885), etc.

Response:

Thank you for directing us to this series of special issues, which contain several relevant studies (the last of these issues was published the day after we submitted our manuscript). In general, the research on the pandemic that is coming out currently may be more thorough compared to some of the initial papers that were published during the pandemic. We agree that including newer studies can strengthen our background section. We have expanded our review of the literature, and have adjusted the manuscript to refer to some of the literature that has come out recently. Among these, we have added references to (Adamus et al., 2022; Airaksinen et al., 2021; Barceló & Sheen, 2020; Duncan et al., 2009; Ebrahimi et al., 2021; Horwood et al., 2023; Panish et al., 2023; Webster et al., 2020; Willroth et al., 2021) in the Introduction. However, our reading of the additional literature appears to mostly supplement what we had included before, and has not lead to any significant changes in our theoretical framework or hypotheses. We have tried to revise section "1.3.2.: How personality may impact pandemic behaviour" to more clearly respond to the reviewer's question "why do personality differences matter". Similarly, section "1.4.: Knowledge gap" has been revised to more directly address the reviewer's questions about how the current work can contribute to the existing literature, in particular in its final sentence:

> *"As the majority of the relevant literature has focused on the association between personality and compliance, we think it could be valuable to also include the association between personality and perceived risk."*

2. The second paragraph in the intro makes the claim that much of the existing work on personality and pandemic behavior was not performed according to current standards of open science - I don't know if this claim can really be substantiated, or if it's even a necessary statement to make. If the authors would like to persist with this type of statement, I recommend framing it more positively, e.g., "adopting more open practices can help clarify or further confirm previous findings" or something along these lines.

Response:

We have moved the relevant paragraph to section "1.4: Knowledge gap". We have reframed the argument somewhat, and we hope the revised version is more aligned with what the reviewer had in mind. This change also aims to address a similar concern expressed by the other reviewer. See the revised text in the response to the other revierer's fifth issue above.

3. Some of the sections describing Big Five personality and pandemic perceptions/behaviors seem a bit repetitive and the writing could be more concise/more organized.

Response:

We recognize that our idea for organizing the Introduction in the previous version of the manuscript (first introducing all of the personality factors and then discussing their relevance to the pandemic) may not have been optimal. Based on this comment, we thoroughly revised the Introduction, and among other things have split up what used to be the long second paragraph in section 1.1.3, and have placed those sentences where each of the factors are discussed in section 1.3.3-1.3.10. This has allowed us to remove some of the redundancies, and hopefully this makes the Introduction easier to read. Our response to the other reviewer's fourth comment should also be relevant here.

4. The authors make the argument that because much of the existing findings rely on cross-sectional studies in which personality is measured simultaneously with pandemic perceptions/behaviors, that these responses can influence each other. The authors' own data collected personality measures 1.5yrs before the pandemic. While this can get around the issue the authors described, it does potentially create a new one: That is, although personality changes are generally shown to be gradual, it may be possible that a "sudden" global event such as a pandemic may lead to more drastic personality changes. As well, other events may have occurred during the two years between personality assessment and pandemic responses data collection that also impact participants' personality. Essentially, is it a fair assumption that personality measures obtained 1.5-2yrs before are still the most accurate in reflecting participants' actual personality characteristics during the pandemic?

Response:

Note that the discussion of potential challenges of cross-sectional studies has now been moved to section "1.4: Knowledge gap" (second and third paragraph), and has been thoroughly revised. Hopefully, the benefits of longitudinal measures for studying this subject are now presented more clearly.

We agree with the reviewer that potential weaknesses of using longitudinal personality measures could also be discussed. We think it would be best to discuss this as a potential limitation of our current study. Thus we have prepared the following text which we plan to include under the heading "Limitations" in the Discussion section of the Stage 2 manuscript:

> *"Using longitudinal data to investigate the relationship between personality traits and pandemic risk perception and compliance may involve some challenges as well, as personality traits do change somewhat over time, a fact that is increasingly recognized among personality researchers (Bleidorn et al., 2021). Further, it could be reasoned that dramatic life events, which some individuals experienced due to the pandemic, are particularly likely to involve personality changes (Rudolph & Zacher, 2023). It is, however, important to note that personality traits have been found to be rather stable, particularly over periods of only a couple of years, and research has suggested that most single, isolated life events result in no or minimal personality change (Bleidorn et al., 2021; Bühler et al., 2024). Thus, the impact of potential personality change on*

*the findings in longitudinal studies on personality traits is likely to be limited,
especially in studies in which personality traits and the outcomes of interest is
measured with only a few years a part."*

*However, in addition to the discussion cited above, we should also note that in our current study
personality measures were taken a relatively short time (the preceding year) before the outcome
measures. Further, the outcome measures were fairly early in the pandemic, when few had yet been
infected in Norway, and life impacts were smaller than they would have been further along into the
pandemic. We therefore do not see the this as invalidating a longitudinal approach to measuring the
effects of personality, and would prefer to retain our description of a longitudinal approach as
preferrable to a cross-esectional one. If the reviewer's point holds that personality may be changed by
dramatic life-events, this may lead our measures of personality to have no or a weak association with
our pandemic outcomes, which we will then have to discuss in the Stage 2 manuscript.*

5. Given that the authors will use the TIPI in their study, it would be good to provide
   additional discussion on personality measurement differences and their impacts on
   related outcomes in their eventual discussion. While I understand the reason for using the
   TIPI, the small number of items in this measure can lead to potential issues with criterion
   validity. Somewhat related, see Bakker & Lelkes (2018):
   https://www.journals.uchicago.edu/doi/full/10.1086/698928.

Response:

*Thank you for bringing the issue of personality measurement differences to our attention, and for the
relevant and interesting reference suggestion! We agree that this issue should be addressed, but we
think it would be most suitable in the Discussion section which will be added to the Stage 2
manuscript. We plan to add something along the lines of the following to the Limitations section:*

> *"Another limitation relates to the TIPI measurement scale used to measure the
> personality traits. TIPI is a short scale including only two items per trait. Although
> short scales have some advantages, they might also affect the obtained results
> as they are unlikely to capture all components of a trait (Bakker & Lelkes, 2018).
> Further, the scale consists of adjectives describing different traits, as opposed to
> statements as many other scales. Although the use of an adjective-based scale
> is not a limitation, it should be noted that it may have affected the result as
> adjectives may lead the respondents to answer more based on how others may
> perceive them as compared to statements which often are more oriented toward
> the respondent's inner experiences (Erevik et al., 2023)."*

6. The authors should make sure to report alpha reliability for all measures used, when
   they arrive at the analysis stage of the report.

Response:

*We plan to report McDonald's omega for internal consistency of the items, and have now specified
this in section "2.4: Analyses". Also see our comment to a similar issue raised by the editor.*

7. It would be good to provide more details on the specific analyses - right now, it looks like there will be two regression analyses reported? Will the authors control for any covariates? Will there be analyses examining potential mediation models, e.g., personality -> perceived risk -> compliance? I assume the authors will also report basic descriptives and zero-order correlations? If so, please include these in the Analyses section.

Response:

We have tried to express the planned analyses more clearly, in the section 2.4 "Analyses" and in Table 2. The reviewer is correct that we are planning two multiple regression analyses. We have added to the third paragraph of section 2.4 that we will include descriptive statistics a correlation table:

> *"We will report descriptive statistics and a correlation table between all seven variables in the model. Bivariate correlations will be reported and investigated in order to describe the data and to test for multicollinearity. Correlations above .7 between independent variables will be interpreted as multicollinearity, as recommended in Pallant (2020)."*

The various contributions of the individual traits will be taken into account by including them all in the same multiple regressions. We do not plan to explicitly control for any additional covariates. We considered more complex modelling, e.g. by planning to use gender or age as covariates (as indicated by Bleidorn et al., 2022), to look for interactions between traits (Musek, 2007), or to test causal relationships between perceived risk and compliance. However, we currently favour keeping the hypotheses relatively simple (although numerous) for the planned analysis. We would like to include exploratory analyses in the Stage 2 manuscript which may investigate any identified relationships in more depth. But we do not feel there is sufficient basis in the literature to state any a priori specific expectations for e.g. covariates or a causal diagrams.

8. The table included does shed some light on the analyses to be conducted - however, I noticed that one set of analyses only included 3 of the 5 traits as predictors (Extraversion, Openness, and Neuroticism), whereas the other analysis includes all five traits. I personally feel that these analyses should be consistent across the board and examine all 5 traits. That is, there is the possibility that the two traits not included in the first set of analysis (Agreeableness and Conscientiousness) could still predict risk perception. The authors should provide more justification for why they set up the regression models as they currently have.

Response:

The newly added Table 1 may make it easier to get an overview of the hypothese. We agree that the relationship between all personality traits and the two outcomes should be investigated. We do plan to test all possible associations between the 5 traits and the 2 outcomes (see cited text at the end of our response). However, we have not made any hypothesis on the potential relationship between agreeableness and conscientiousness and risk perception as we have not found theoretical or empirical support for such hypotheses. This is argued for in section 1.3.3-1.3.10. Also note that the other reviewer appears to argue that we may have already included "too many" hypotheses. We have added explicit statements about the non-hypothesized associations at the end of 1.3.3:

> *"To our knowledge, there is no empirical or theoretical reason to expect an association between conscientiousness and risk perception."*

And at the end of 1.3.4:

> *"To our knowledge, there is no empirical or theoretical reason to expect an association between Agreeableness and Risk perception."*

If the two non-hypothesized associations should be found to be significant, we will report this as a novel discovery, rather than as a supported hypothesis. We have now sought to state this plan more clearly by adding the following text to the first paragraph of section 2.4:

> *"We will use one multiple linear regression to test potential all ten relationships between the five personality traits and Perceived risk. We will use a second multiple linear regression to test potential relationships between the five personality traits and Compliance. the two pandemic outcomes using two multiple linear regressions (one for Perceived risk and one for Compliance). As can be seen in Table 1 and Table 2, there are hypotheses associated with eight of the ten possible combinations in these analyses. Any effects on the two non-hypothesized associations will be reported as novel discoveries."*