**Reviewer Comments**

*Revisiting and updating the risk-benefit link: Replication of Fischhoff et al. (1978) with*

*extensions examining pandemic related factors*

**Overall Comments**

**I enjoyed reading this detailed and well-written document. This registered report provides a good account of detailing a replication attempt. I believe that Fischoff et al. (1978) is a suitable study to replicate, and the addition of COVID-related items provides an interesting modern context. I have identified several areas which I feel require extensive revision to clarify to the reader 1) why a replication is needed at this time, 2) how will this replication provide new relevance in addition to that conducted in 2016, and 3) why there are considerable deviations from the original study protocols. You only use 14/30 of the original study items. These deviations from the original, and subsequent replication, protocols make me think that it is important to highlight to the reader that this should be considered a partial, or perhaps conceptual replication of the original study.**

 **I recommend that the authors clarify the following points before further consideration.**

**Stage 1 Submission Requirements:**

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

**As this is a replication, I will refer to the validity of the replication attempt and its current relevance.**

***Choice of study for replication: Fischhoff et al. (1978) section.***

**This section is too long and does not seem appropriate for scholarly publication. It gives the sense that you have chosen to replicate this study mainly because it is popular. For example, discussing google scholar citations, and contacting the authors, may be suitable for a graduate thesis, but not a published article. To justify the need for this replication study, greater attention should be given to highlighting why replicating historical findings is important (expand on the final sentence of this section), and why this specific replication is relevant now. The question of timing is particularly important given that you highlight the recent replication in 2016. Is an increased sample your main contribution? Are there COVID-related contextual factors that you think are important to capture with respect to the original findings?**

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

***Methods section***

**You should justify why you expect to find support the original negative association between perceived benefit and perceived risk ratings, given that Fox-Glassman and Weber (2016) failed to find support. Why do you think the former and not the latter will be repeated?**

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

***Power and sensitivity analyses section.***

**This section is confusing. “*We aimed for a sample of 1000 participants, 333/4 in each condition, which a sensitivity analysis indicated would allow the detection of independent samples t-test of the two conditions*” This first suggests 3 conditions (1000/3) then states 2 conditions. You later state “*Participants are randomly assigned to either Task 1a, Task 1b or Task 1c*” suggesting 3 conditions. Please clarify and word this more clearly. Also, your stage 1 snapshot states you are going for .95 power not .8 – make sure these are consistent.**

**Given that you have stated that the core of this replication is the relationship between perceived risk and perceived benefit, this should serve as the basis for your main power analysis. Highlight this. Also, why is this based on correlations of .14? This is not a fatal point, just specifying this value needs explaining.**

*Participants section.*

Instead of listing CloudResearch options and Qualtrics functions (which will be a little value to the reader), it is important that you explain the criteria for participation of your target population. Can anyone participate? Will it be a representative sample? Are there any relevant restrictions on participation? Will there be any criteria for quotas/exclusion of participants?

With respect to who your sample will be, and the extent to which they are representative of the broader population, you should refer back to the sampling of Fischoff. You have highlighted that the original study was too small in sample size. Additionally, it was a sample of a specific group of people affiliated with a political group known for enacting societal change. In my opinion, this makes it highly likely that their opinions may not be representative. This may explain why the original study sample broadly reported that ‘serious action’ was needed to mitigate the level of most risks, whereas this was not replicated in the 2016 study. Again, further explanation is needed to fully explain the sizeable jump in sample (beyond what may normally be expected in replication attempts).

In terms of remuneration, this amount appears to be well below half of what might be considered a living wage hourly rate in the US. Are there any ethical concerns around valuing participants time?

**I got the sense that the dramatic reduction in study duration was due to the fact you had already chosen the online surveying platform. Therefore the choice of method (MTurk) seems to supersede the replication methodology (justification for dramatically reducing study duration from 90-120 minutes to 10-20 minutes). I understand the practicalities of this in terms of funding etc, so I’m not suggesting you change this. Perhaps try to word it more that there is a trade-off between needing to dramatically increase the sample to identify whether there is really an effect going on, but that the cost of this is that you need to alter the protocol to reduce time.**

There are also changes to the original protocol such as the original had judgements on a 10 point scale, yours in 1000. Is there a justification for this?

***Extensions section.***

**Of the 4 listed items relating to the pandemic, I struggle to see the direct relevance of your inclusion of ‘Biological weapons’ and would suggest more obvious activities relevant to mitigating the risks of the pandemic like mask wearing.**

In the analysis section in the original study, they highlight their rationale for using geometric not arithmetic means. Is there a justification for your deviation from the original analysis?

**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.**

**See methodological point in section 1C above.**

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

*Evaluation criteria for replication findings section.*

This section needs elaboration. Which specific statistical tests, with which specific variables, which directions of effect etc? These need to be stated much more clearly so there can be no doubt as to what will and will not constitute a successful replication.

**Additional suggestions and comments**

***Introduction section.***

**When introducing the work of Fischoff at the beginning, it might be useful to also briefly discuss the work of Chauncey Starr to highlight where this study sits in the context of the historical literature and the difference between ‘revealed preferences’ and ‘expressed preferences’.**

**You state the “*goal was to conduct an independent replication of the negative correlation*” - the goal should be to replicate the study protocols, not specifically to reproduce a result.**

**Sometimes you refer to the relationship between perceived risk and perceived benefit, sometimes just to the risk/benefit relationship. It is important to be consistent and highlight to the reader that it is the perceived, not objective, relationship that is being studied.**

***Pre-registration and open-science section.***

**You don’t need a section for this, similar to my comment about the choice of study replication section, this reads more like a student thesis than publishable article, referring the reader to the osf link is sufficient.**

Additional suggestions

You may wish to consider the influence of individual level numeracy. Recent research has highlighted that people can struggle to numerically express their beliefs and perceptions of the level of risk they experience. This is particularly true when trying to use large numbers (for instance providing a score out of 1000). For example, Raude et al. (2021) recently reported that the magnitude of the primary bias in risk (overestimating uncommon risks and underestimating common risks) varies as a function of the respondents’ individual level of numeracy. Given this recent finding that a famous effect within the risk perception literature is heavily influenced by numeracy, you may consider adding a short measure of numeracy to consider whether the Fischoff effect is also partially driven by this factor.

Additionally, having looked at your Qualtrics survey, you may wish to consider randomising the order of the 18 items to avoid any order effects. These are common to occur in questions with large numbers of sliding scales.

You may also look to highlight, either in the existing introduction or in the later discussion, some of the developments in understanding risk perceptions since 1978 that may influence our interpretation of the results. For example, my colleagues and I continue to explore the environmental and informational cues of risk perceptions, which can often differ depending on the demographic characteristics of the sample. Given your extension of the replication to COVID behaviours, see our recent work on risk perceptions during the pandemic and assessment of protective behaviours <https://link.springer.com/article/10.1007/s10389-021-01543-9> & <https://www.tandfonline.com/eprint/BSGYMZUI79CSNGD9VGRH/full?target=10.1080/13669877.2021.1908403>.

All the best,

Richard Brown,

Psychology Department,

Northumbria University,

Newcastle, United Kingdom