## Revisiting and updating the risk-benefit link: Replication of Fischhoff et al. (1978) with extensions examining pandemic related factors

Jason M. Frank  
ORCID: 0000-0001-5483-1991  
Department of Psychology, University of Hong Kong, Hong Kong SAR  
[jfrank@connect.hku.hk](mailto:jfrank@connect.hku.hk) / [jf7082@gmail.com](mailto:jf7082@gmail.com)

^Gilad Feldman

OCRID: 0000-0003-2812-6599  
Department of Psychology, University of Hong Kong, Hong Kong SAR  
[gfeldman@hku.hk](mailto:gfeldman@hku.hk) / [giladfel@gmail.com](mailto:giladfel@gmail.com)

^Corresponding author

## Author bios:

Jason M. Frank was a guided thesis master’s student at the University of Hong Kong during the academic year 2021-22.

Gilad Feldman is an assistant professor with the University of Hong Kong psychology department. His research focuses on judgment and decision-making.

## Declaration of Conflict of Interest:

The author(s) declared no potential conflicts of interests with respect to the authorship and/orpublication of this article.

## Financial disclosure/funding:

The author(s) received no financial support for the research and/or authorship of this article.

## Authorship declaration:

Jason M. Frank conducted the replication as part of Dissertation in Psychology PSYC7308 course.

Gilad advised and supervised, conducted the pre-registrations, and ran data collection.

## Corresponding author

Gilad Feldman, Department of Psychology, University of Hong Kong, Hong Kong SAR; [gfeldman@hku.hk](mailto:gfeldman@hku.hk) ; 0000-0003-2812-6599

## Rights:

CC BY or equivalent license is applied to the AAM arising from this submission. ([clarification](https://bit.ly/rrs-primer))

## Important links and information

Citation of the target research articles:

* Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., & Combs, B. (1978). *How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. Policy Sciences, 9(2), 127–152*. <https://doi.org/10.1007/BF00143739>
* Fox-Glassman, K. T., & Weber, E. U. (2016). What makes risk acceptable? Revisiting the 1978 psychological dimensions of perceptions of technological risks. *Journal of Mathematical Psychology, 75,* 157 169.<https://doi.org/10.1016/j.jmp.2016.05.003>

## Contributor Roles Taxonomy

The table below employs CRediT (Contributor Roles Taxonomy) to identify the contribution and roles played by the contributors in the current replication effort. Please refer to the url (<https://www.casrai.org/credit.html> ) on details and definitions of each of the roles listed below.

|  |  |  |
| --- | --- | --- |
| **Role** | **Jason M. Frank** | **Gilad Feldman** |
| Conceptualization | X | X |
| Pre-registration | X |  |
| Data curation |  | X |
| Formal analysis | X |  |
| Funding acquisition |  | X |
| Investigation | X |  |
| Pre-registration peer review / verification |  | X |
| Data analysis peer review / verification |  | X |
| Methodology | X |  |
| Project administration |  | X |
| Resources |  | X |
| Software | X |  |
| Supervision |  | X |
| Validation |  | X |
| Visualization | X |  |
| Writing-original draft | X |  |
| Writing-review and editing |  | X |

# Abstract

**[IMPORTANT:   
Method and results sections were written using a randomized dataset produced by Qualtrics to simulate what these sections will look like after data collection. These will be updated following the data collection. This is written in past tense yet no pre-registration or data collection have been conducted.]**

The relationship between risk and benefit is complex and has been studied in a number of different ways. Fischhoff et al. (1978) approached the relationship in a novel way, using psychometric analysis to measure public attitudes regarding the perceived risks and perceived benefits of certain technologies and activities, finding a negative correlation between perceived risk and perceived benefit. In a well-powered, pre-registered study we set out to conduct a replication of Fischhoff et al. (1978) with several adjustments and extensions. Using a simplified survey design with dummy data generated in Qualtrics (*N* = 1000) and improved statistical testing, we conducted an exploratory analysis indicating that participants [did/did not] rate risks and benefits differently with respect to the selected technologies and activities. Extending the study, we added (1) items related to the ongoing Covid-19 pandemic to a modified list of technologies and activities from the original for which, using dummy data generated in Qualtrics (*N* = 1000), we found [weak to no] empirical support for the negative relationship between perceived risk and perceived benefit and (2) a within subjects condition to the study, which we will analyze for empirical support for the negative correlation between perceived risk and perceived benefit. Supplementary, materials, data, and analysis files/code are available here: <https://osf.io/hcvmz/>.

*Keywords:* judgment and decision-making, risk, perceived risk, perceived benefit, psychometric risk, psychological risk dimensions, risk acceptability, replication

# PCIRR-Study Design Table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Question | Hypothesis | Analysis plan | Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis | Interpretation given different outcomes | Theory that could be shown wrong by the outcomes |
| What is the relationship between perceived risk and perceived benefit? | Negative correlation between perceived risk and perceived benefit | Independent samples t-tests  Correlation / Linear regression | Following analysis in original article and updating to improve analyses and reporting of results | We examine the replicability of Fischhoff et al. (1978) and support for our suggested extensions | Negative correlation between perceived risk perceived and benefit |
| What is the relationship between acceptable levels of risk judgments and perceived risk and perceived benefit? | Exploratory | Independent samples t-tests  Correlation / Linear regression | N/A |
| What is the relationship between characteristics of risk and perceived risk and benefit? | Exploratory | Independent samples t-tests  Correlation / Linear regression | N/A |
|  | | | | | |

# Revisiting and updating the risk-benefit link: Replication of Fischhoff et al. (1978) with extensions examining pandemic related factors

## Background

An economic analysis of the relationship between risk and benefit would proceed on the standard assumption of a trade-off between risk and return. Such an analysis would thus generally predict a positive correlation between risk and benefit – as the risks of an activity or technology increases, its benefits must also increase if that activity or technology is to be accepted by society. In people’s minds, however, this does not appear to be the case. Fischhoff et al. (1978) was an early attempt to study the relationship between risk and benefit as perceived by people, as opposed to observed in the world, and demonstrated a negative relationship between perceived risk and perceived benefit.

We aimed to build on the work by Fischhoff et al. (1978) with several goals. Our first goal was to conduct an independent replication of the study using a simplified and scalable psychometric paradigm. Our second goal was to extend the application of the psychometric paradigm to examine the risk/benefit relationship for a limited set of Covid-19 pandemic-related activities and technologies.

We begin by introducing the literature on psychometric risk and benefit assessment, their negative correlation, and the chosen article for replication - Fischhoff et al. (1978). We then discuss our motivation for the current study, provide an overview of the findings from the original study, and finally present our adjustments and extensions to the original’s design. Due to the number of deviations from Fischhoff et al. (1978) we categorized this study as falling between a close and far (i.e., conceptual) replication according to the criteria set forth in LeBel et al., (2018).

## The relationship between risk and benefit

All human activity entails the potential for risk, be it the use of home appliances, sourcing electricity from nuclear power plants, or deciding whether to be vaccinated against the Covid-19 virus. Standard economic analysis dictates that the optimal way to determine the acceptability of an activity or technology would be to analyze its risks and benefits to determine the best trade-off (Sharpe, 1964). Accordingly, such an approach would predict that if a very risky activity or technology were to be accepted by individuals or society as a whole, its benefits would have to be correspondingly high. On the other hand, risk perception – that is, how individuals intuitively perceive and judge the riskiness of things and whether to accept them, seems to work quite differently.

Unlike economists and technological experts, most people seem to rely on their intuitive risk perceptions to judge and evaluate activities and technologies. These intuitions are likely informed by their past experiences, including exposure to the relevant activity or technology and media about it and mishaps connected to it.. Fischhoff et al. (1978) was among the first to study these intuitions in a measurable, quantifiable way by using a psychometric questionnaire. Participants were asked to make quantified judgments about certain activities and technologies with respect to such items’ current and desired benefits and risks as well as other characteristics such as their newness and controllability. These judgments were then used to explore relationships between perceived risk and perceived benefit and other factors such as acceptable levels of risk and the various characteristics of risk. One of the core findings of Fischhoff et al. (1978) was that contrary to the prediction of economists, people tend to view risk and benefit as negatively correlated. For example, people typically perceive smoking and alcoholic beverages as being very high in risk while being very low in benefit and conversely that antibiotics, vaccines, and X-rays are very high in benefit but very low in risk (Fischhoff et al., 1978). In people’s minds, it appears to be that the greater the perceived benefit of an activity or technology, the lower the perceived risk and vice versa (Fischhoff et al., 1978, McDaniels et al., 1997).

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

We chose the Fischhoff et al. (1978) study based on several factors: impact, absence of well-powered, pre-registered replications, potential for improvement on methodology and statistical analysis, and potential for extensions relating to the current ongoing COVID-19 pandemic.

Fishhoff et al. (1978) has had a significant impact on scholarly research in the area of behavioral economics and judgment and decision making, most notably by introducing a psychometric paradigm to the study of the risk and benefit relationship. Many important follow-up theoretical and empirical articles have been published, such as Slovic (1987) summarizing Fischoff et al. (1978) and related follow-up research, a line of research on the affect heuristic in risk perception (Finucane et al., 2000; Slovic et al., 2007), as well as a large body of research using the factors of dread and unknown risk to understand risk perception (Gigerenzer, 2004; Koonce et al., 2005). Fishhoff et al. (1978) has also produced impactful research regarding communication and education of the public regarding risk (Slovic, 1986) and the social amplification of risk (Kasperson et al., 1988). At the time of writing, there were 3,814 Google Scholar citations of the article. Following the growing recognition of the importance of reproducibility and replicability in psychological science (e.g., Brandt et al., 2014; Open Science Collaboration, 2015; van‘t Veer & Giner-Sorolla, 2016; Zwaan et al., 2018), we believed that an impactful study such as Fischhoff et al. (1978) was a strong candidate for replication.

While Fox-Glassman and Weber (2016) conducted a replication of Fischhoff et al. (1978), that study was not pre-registered and had mixed replication success. In particular, it was unable to replicate the negative correlation between perceived risk and perceived benefit. More importantly, Fox-Glassman and Weber (2016) and the original study both were likely underpowered, each with only 75 and 76 participants, respectively. Accordingly, we believed that Fischhoff et al. (1978) remained a relevant target for replication, in particular with a much larger sample size. Moreover, as we began constructing the replication design to accommodate a larger sample size, we realized a number of areas for improvement relating to the duration of the study, the methodological approach, as well as the statistical analysis in the original, each of which are explained in detail below. Finally, given the original study’s ability to inform us about people’s risk perception, we believed it would be informative to apply it to the COVID-19 pandemic. Given these reasons, we aimed to revisit this classic study and embarked on a well-powered replication of the classic perceived risk-benefits link with an improved and extended design of Fischhoff et al. (1978).

## Overview of the replication and extensions

### Replication

Fischhoff et al. (1978) was based on three main tasks requiring participants to judge and quantify 30 technologies and activities with respect to: (1) the perceived benefit or risk of each item, (2) how acceptable each item’s current level of risk is, and (3) each item’s rating on nine characteristics of risk scales. Fischhoff et al. (1978) did not have explicit hypotheses relating to its data and analyses, yet reported many findings. One key finding was the negative correlation between perceived risk and perceived benefit, indicating that people’s intuitions about risk and benefit trade-offs do not follow a standard economic formula. Fischhoff et al. (1978) also reported a number of findings correlating risk acceptability scores and level of acceptable risk scores (as explained more fully below) with perceived risk and perceived benefit. Finally, using the nine characteristics of risk ratings, Fischhoff et al. (1978) reported that these characteristics reduced to two main factors: “dread” and “unknown” and further that together with perceived benefit, these two factors correlated with perceived risk. As already noted, Fox-Glassman and Weber (2016) reported mixed success with replication of Fischhoff et al (1978)’s findings. In particular, Fox-Glassman and Weber (2016) failed to find support for the negative correlation between perceived risk and perceived benefit.

While we retained all three tasks for this replication, we made major adjustments to create a much simplified questionnaire and shorter duration as well as other significant modifications to address the various areas for improvement noted above. With respect to duration, both studies had lengthy durations making replication with a large sample size difficult. Fischhoff et al. (1978) reported an average duration per participant of two hours and Fox-Glassman and Weber (2016) reported a mean duration per participant of 90 minutes. Both seem to be very taxing to the point of being extremely difficult for participants, especially so for online labor market workers, as in the case of the sample used by Fox-Glassman and Weber (2016). Our experience with Amazon Mechanical Turk (MTurk) showed fast decreasing attentiveness and fast increasing attrition rates for studies longer than 10-20 minutes, regardless of pay compensation. Going beyond duration, we found the methodology used in the original to be somewhat difficult and confusing for participants to understand and process well, and thought it necessary to make some methodological adjustments. We therefore decided on a modified psychometric paradigm. Accordingly, using the materials provided by the authors of Fox-Glassman and Weber (2016), we constructed a simplified survey suitable for a much larger sample size, which we believe allows for higher quality responding and potentially more accurate insights and conclusions about the relationship between perceived risk and perceived benefit. We provided a full list of deviations and explanations for the deviations between the original study and Fox-Glassman and Weber (2016) and the present replication in Table 3 in the supplementary materials.

Secondly, and more importantly, we re-examined the data analysis strategy used in both studies and identified many areas for improvement. Fischhoff et al. (1978) recognized that the basic structure of its survey was an issue in terms of duration for participants. To mitigate this, the authors adopted a between subjects design, splitting participants into two conditions, one rating only perceived benefit, the other rating only perceived risk. For the core part of their analyses, Fischhoff et al. (1978) and Fox-Glassman and Weber (2016) then used item-level mean ratings to correlate and regress results across these two conditions. However, due to the small number of items used in both studies, the ability to detect significance in the relationship between ratings differences on an item-level would require an extremely large and somewhat unlikely effect given common correlations in social psychology. We believe this may explain the mixed results present in both studies. To be able to address the research question we would require either many more items, or an analysis on a participant rather than an item level. To improve the study, we modified the analysis of the two conditions to instead perform the only participant level analysis suitable for this design: an independent samples t-tests comparing the participant-level ratings for each item. We believe this provides more accurate and reliable results with respect to the differences between the perceived risk and perceived benefit ratings. We note that this analysis is primarily exploratory for purposes of this study as it will not test the theoretical relationship between perceived risk and perceived benefit as contemplated by Fischhoff et al. (1978) and subsequent studies. However, more importantly and more relevant for the testing of the benefits-risks link, we added a third condition, detailed below as Extension 2, displaying both risks and benefits to participants, which allows us to examine the risk-benefit associations with sufficient power.

In our replication we focused primarily on the negative relationship between perceived risks and perceived benefits. This relationship has been demonstrated in numerous studies since Fischhoff et al. (1978) (Alhakami and Slovic, 1994; Finucane et al., 2000; McDaniels et al., 1997; Skagerlund et al., 2020; Slovic et al., 1987), most recently in a replication of Finucane et al. (2000) conducted two decades after the original with samples from the US and the UK (Efendić et al., 2021). Accordingly, we expect results to show support for the negative correlation between perceived risk and perceived benefit. Our main test for this hypothesis is by examining participant-level risk-benefit associations in an extension, explained in detail in the section “Joint risks-benefits condition” below. In addition, to make the most of the replicated design we will also be conducting independent samples t-tests examining differences in participants’ perceived risk and perceived benefit ratings. This may begin to map whether certain types of technologies or activities are perceived differently by participants. We intended this analysis as exploratory, and we suggest caution in the possible interpretation of these findings. We hope that this adaptation of the target’s design can help motivate a future line of research that would further compare risks and benefits of technologies and aim to make sense of how participants process the two. This would also allow for an exploratory comparison of the within-subject and between-subject designs, given that these are combined into a unified design. In the supplementary materials, we summarized the key findings in Fischhoff et al. (1978) and Fox-Glassman and Weber (2016) in Table 2 and our deviations from the original and Fox-Glassman and Weber (2016) in Table 3.

### Extensions

#### Extension 1: Pandemic related items (exploratory)

Building on the original’s design, we added several items that relate to the ongoing Covid-19 pandemic, including: 1) Covid vaccinations, 2) Experimentation with biological viruses, 3) Lockdowns to address Covid-19 pandemic, and 4) Social distancing to address Covid-19 pandemic. The aim of this extension was to gain insights as to people’s evaluations and judgments concerning the benefits and risks of various pandemic responses and policies. In particular, we will explore whether participants view the relationship between the perceived risks and perceived benefits of these items differently than other non-pandemic related activities and technologies, and the differences between perceived risks and benefits. If participants do view the relationship differently, this may provide useful insights as to how to structure pandemic related public communications around the pandemic, especially regarding activities and technologies designed to mitigate the pandemic or protect the public. Indeed, the relationship between COVID-19 risk perception has been associated with adherence to pandemic prevention measures (Brown and Pepper, 2021) and further insight may be instructive. Measurements and data analysis concerning the additional extension items will be consistent with the main analysis in the study.

#### Extension 2: Joint risks-benefits condition

The third condition (Task 1c explained in detail below) will ask participants to rate both perceived risk and perceived benefit, thereby allowing for testing of correlation between perceived risk and perceived benefit ratings at the participant level as opposed to the item-level. This is an improvement to the design of the original study as it will provide the test needed to address the core hypothesis underlying the original study: the relationship between perceived risks and benefits. We expect this condition to show a negative correlation between perceived risk and perceived benefit consistent with numerous studies since Fischhoff et al. (1978) (Alhakami and Slovic, 1994; Efendić et al., 2021; Finucane et al., 2000; McDaniels et al., 1997; Skagerlund et al., 2020; Slovic et al., 1987).

## Pre-registration and open-science

We will pre-register the experiment on the Open Science Framework (OSF) and data collection will be launched shortly after pre-registration. Pre-registrations and all materials used in these experiments are available in the supplementary materials. We provided all materials, data, code, and pre-registration on: <https://osf.io/hcvmz/>.

We provided additional open-science details and disclosures in the supplementary materials under “Open Science disclosures” sub-section. All measures, manipulations, exclusions conducted for this investigation will be reported, all studies will be pre-registered with power analyses, and data collection will be completed before analyses.

# Method

**[IMPORTANT:   
We provided partial results in the following sections written using a randomized dataset produced by Qualtrics to simulate what these sections will look like after data collection. These will be updated following the data collection. No pre-registration or data collection took place yet.]**

## Power and sensitivity analyses

Given the adjustments made to the original’s design, the original’s focus on item level analyses, and our data analysis plan to focus on individual level, we felt that the original’s effect cannot serve as a basis for the replication power analysis.

We aimed for a sample of 1000 participants, to be evenly split among our three conditions (Task 1a, Task 1b and Task 1, as explained below), resulting in 333/4 in each condition. As explained more fully below, a data analysis strategy contemplates conducting independent samples t-tests on the results from Task 1a and Task 1b. A sensitivity analysis indicated that a sample size of 333 participants in each of these two conditions would allow the detection of independent samples t-test with an effect of *d* = 0.32 (given 333 in each condition for two condition comparisons, power = 80%, alpha = 0.1%, two-tailed), traditionally considered a medium effect. Separately, for Task 1c, we will be conducting a correlations analysis. A sensitivity analysis indicated that would allow us to detect correlations of *r* = 0.15 (given single condition of 333, power = 80%, alpha = 5%, two-tailed), traditionally considered a small to medium effect. Following data collection, we will provide an updated sensitivity analysis for any reduction due to exclusions. A 10% reduction in the number of participants would result in ~33 fewer participants per condition, meaning ~300 per condition. This would still capture an effect of .34 for independent samples t-test and .16 for correlation, which we believe would be sufficient for the stated purposes in the study. Our planned sample is several times larger than both Fischhoff et al. (1978) and Fox-Glassman and Weber (2016) that had 75-6 participants.

## Participants

To demonstrate what the results would look like after data collection we simulated a dataset of 1000 participants using Qualtrics and report our analyses below based on that dataset. Results will later be updated in full to a sample of 1000 and the real data.

We will recruit participants from Amazon Mechanical Turk using the CloudResearch/Turkprime platform (Litman, Robinson, & Abberbock, 2017). We will define the HIT for participants that (1) have a HIT Approval Rate between 95% and 100%; (2) have between 5,000 and 1,000,000 tasks, and (3) are located in the United States. In addition, based on our extensive experience of running similar judgment and decision making replications on MTurk, to ensure high quality data collection, we will employ the following CloudResearch options: Duplicate IP Block. Duplicate Geocode Block, Suspicious Geocode Block, Verify Worker Country Location, Enhanced Privacy, CloudResearch Approved Participants, Block Low Quality Participants, etc. We will also employ the Qualtrics fraud and spam prevention measures: reCAPTCHA, prevent multiple submission, prevent ballotstuffing, bot detection, security scan monitor, relevantID, etc.

Assignment pay is based on the federal wage of 7.25USD/hour, per minute, so for example - 5-8 minutes survey would be paid 1 USD per participant. [We will first pretest survey duration with 30 participants to make sure our time run estimate was accurate and adjusted pay as needed, the data of the 30 participants will not be analyzed other than to assess survey completion duration and needed pay adjustments. For those pretest participants, if the survey duration is longer than expected, they will be paid a bonus as pay adjustment.]

As noted in Fox-Glassman and Weber (2016), Fischhoff et al. (1978) did not provide a breakdown of its sample population’s demographics. However, it did note that the participants were all members of the Oregon League of Women Voters, which was described in Fischhoff et al. (1978) as “a generally liberal, environmentally minded group”. Accordingly, in addition to allowing an increased sample size, the use of Amazon Mechanical Turk offers the potential for a more diverse sample population than the original study.

## Table 1 *Difference and similarities between samples from Fischhoff et al. (1978), Fox-Glassman and Weber (2016), and the current study*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Fischhoff et al. (1978)  Eugene, Oregon, League of Women Voters | Fox-Glassman and Weber (2016)  US MTurk workers | Replication and extension  US MTurk workers |
| Total sample size | 76 | 75 (83 prior to omissions) | 1000 | |
| Perceived benefit | Not reported (~half of total sample size | 33 | ~333 | |
| Perceived risk | Not reported (~half of total sample size | 42 | ~333 | |
| Perceived benefit and risk | Not conducted1 | Not conducted1 | ~3331 | |
| Risk adjustment factor | 76 | 75 | ~666 | |
| 9 characteristics of risk | 76 | 75 | 1000, randomly assigned to 2 out of 9 characteristics (~222 per characteristic)1 | |
| Duration for completion | 2 hours average | 90 minute mean competition time | ~ 20 minutes | |
| Geographic origin | Eugene, Oregon | US American | US American | |
| Gender | 24 males, 52 females, | 37 males, 36 females, 2 unspecified | [XXX] males, [XXX] females | |
| Median age (years) | Not reported | Not reported | [XX] | |
| Average age (years) | Not reported | 36 | [XX.X] | |
| Standard deviation age (years) | Not reported | 12.5 | [XX.XX] | |
| Age range (years) | Not reported | Not reported | [XX-XX] | |
| Medium (location) | Paper-based | Computer (online) | Computer (online) | |
| Compensation | Not reported | Computer (online) | Nominal payment | |
| Year | 1978 | 2013 (data collection), 2016 (publication | 2022 | |
| *Notes*. 1. See Table 3 in the supplementary for full details of deviations in study design from Fischhoff et al. (1978). | | | | |

## Design and procedure: Current study

We summarized our experimental design in Table 2 below, a design based on the three main tasks from Fischhoff et al. (1978). Participants access the online survey via Qualtrics. They first complete a consent form, then read a brief overview of the experiments, followed by verification checks. Next, participants evaluate 18 technologies and activities with respect to: (1) the perceived benefit or risk of each item, (2) how acceptable each item’s current level of risk is, and (3) each item’s rating on 9 characteristics of risk scales.

The list of 18 items was based directly on Fischhoff et al. (1978) but reduced to 14 items to reduce overall study duration and cognitive burden. The subset of items was selected based on various criteria including, relevance to current society, relevance to a broader population, repetitiveness, and clarity. For instance, we found that items related to transportation were overrepresented and deleted “bicycles”, “commercial (private) aviation”, “motorcycles”, and “railroads”, while retaining “general aviation” and “motor vehicles”. Similarly, a number of items were relevant only to a smaller or limited population due to geographical requirements or other reasons. For instance, we deleted “high-school and college football”, “hunting”, “mountain climbing”, “power mowers”, “skiing”, and “swimming”. In the supplementary materials, we provided the full list of items used in Fischhoff et al. (1978) in Table 4 and the list of deleted items and rationale for deletion in Table 5. In addition to the 14 items taken from the original study, we added four Covid-19-related items to the list as an extension: COVID-19 vaccines, experimentation with biological viruses, lockdowns to address the COVID-19 pandemic, and social distancing to address the COVID-19 pandemic. We provided the full list of items used in the current study in Table 5, and summarized the deviations from Fischhoff et al. (1978) and Fox-Glassman and Weber (2016) in Table 3 in the supplementary materials.

Fischhoff et al. (1978) did not specify the order in which the 30 items were presented to participants. In order to control for the potential impact of ordering effects, Fox-Glassman and Weber (2016) randomized the order of presentation. In the current study, we grouped items together based on similarity and presented them uniformly across all three tasks of the study. For instance, we grouped together “nuclear power” and “electric power”, “motor vehicles” and “general aviation”, and “contraceptives”, “prescription antibiotics”, “surgery”, and “X-rays”. While this may create the potential for the impact of ordering effects, we believe this is an improvement in the study design as it should significantly reduce cognitive burden when participants are moving from Task 1 through Task 3 to deal with the same 18 items across different scales.

## Table 2

## *Replication and extension: Survey procedure*

|  |  |
| --- | --- |
| **Task** | **Description** |
| **Task 1** - Perceived benefit and risk rating | Participants are randomly assigned to either Task 1a, Task 1b or Task 1c (Extension 2) |
| **Task 1a** - Perceived benefit | Participants rate each of 18 items in terms of their benefits, ignoring all costs and risks, on a scale from 0 to 1000. |
| **Task 1b** - Perceived risk | Participants rate each of 18 items in terms of their risks, ignoring all benefits, on a scale from 0 to 1000. |
| **Task 1c (Extension 2)** -Perceived benefit and perceived risk | Participants rate each of the 18 items in terms of both their risks and benefits, on a scale from 0 to 1000 |
| **Task 2** - Risk adjustment factor (only participants from Tasks 1a/1b) | 1. Participants rate the current levels of risk for each of 18 items as either :  * “Could be much riskier” * “Could be riskier” * “Are currently acceptable” * “Too risky” * “Much too risky”  1. If any items are rated as “Could be much riskier” or “Could be riskier”, participants are prompted to enter a multiplier to indicate how many more times risky the item could be and still be acceptable. 2. If any items are rated as “Too risky” or “Much too risky”, participants are prompted to enter a multiplier to indicate how many more times safe the item should be to be acceptable. |
| **Task 3** - Characteristics of risk (all participants) | Participants randomly assigned to rate all 18 items 2 out of 9 seven-point scales:   * Voluntariness * Immediacy * Knowledge about risk * Scientific knowledge about risk * Controllability * Newness * Chronic vs. catastrophic * Common vs dread * Severity of consequences |

### Task 1a/1b (replication): Perceived benefit and risk (between-subjects)

We randomly assign participants into one of three conditions for Task 1. Consistent with Fischhoff et al. (1978), Task 1a and Task 1b will ask participants to rate either the benefits or the risks, respectively, of all 18 items in the study. In Task 1a, the first condition, participants are asked to judge the benefits associated with each of the 18 items. These benefit raters are instructed to consider “all types of benefits” associated with each item and not to consider any costs or risks associated with it. In Task 1b, condition two, participants are asked to judge the risks associated with each of the 18 items. These risk raters are instructed to consider “any risk of dying or increased likelihood of dying” as a consequence of each item and not to consider any benefits associated with it. In each condition, participants answer two comprehension check questions to confirm their understanding of the instructions. Participants are then asked to rate each item on a slider scale from zero to 1000, with zero meaning that the item has zero benefit or risk, as applicable, and 1000 meaning it has the maximum benefit or risk. Participants are also instructed to keep the ratings consistent so as to reflect the relative ranking of the items to each other.

**Task 1c (Extension 2) - Perceived benefit and risk (within subjects)**

Task 1c, the third condition, asks participants to rate each of the 18 items on both perceived risk and perceived benefit. The ordering of risk and benefit ratings will be counterbalanced by randomizing which one is presented first to participants. Participants in this within subjects condition will not proceed to Task 2 and will proceed directly to Task 3 in order to keep duration across participants relatively consistent.

### Task 2: Risk adjustment factor

After completing Task 1, only participants that perform Task 1a or Task 1b are asked to complete Task 2. In Task 2 participants are instructed to judge how acceptable the risk level of each item currently is. The instructions clarify that every item carries risk, but that ideally risk should be zero. An acceptable level of risk is described as one that is “good enough” such that the “advantages of increased safety are not worth the costs of reducing risk”. For each of the 18 items participants are asked to rate the current levels of risk as one of the following: “Could be much riskier”, “Could be riskier”, “Are currently acceptable”, “Too risky”, “Much too risky”.

After completing the ratings for all 18 items, participants are then presented with a follow up question. For any item rated as “Could be much riskier” or “Could be riskier”, participants are asked to enter a numerical value to represent how many more times risky such item could be riskier and still be acceptable. Similarly, for any item rated as “Too risky” or “Much too risky”, participants are asked to enter numerical value to represent how many more times safe such item would have to be in order to be acceptable.

Based on participants’ responses to the risk acceptability question and the corresponding multiplier questions, a risk adjustment factor score (RAF score) is calculated. Items rated as “Too risky” or “Much too risky” are given RAF scores equal to responses to the question “X would be acceptable if it were \_\_\_\_\_\_\_ times riskier.” Items rated as “Are currently acceptable” are given RAF scores of 1.0. Finally, items rated as either “Could be much riskier” or “Could be risker” are given RAF scores equal to the multiplicative reciprocal of the response to the question “To be acceptable, X would have to be \_\_\_\_\_\_\_ times safer.” RAF scores greater than 1.0 indicate that current risk of the activity or technology is too high and must be lowered. RAF scores of 1.0 indicate that the current risk is acceptable and RAF scores less than 1.0 mean that the risk could be higher.

Similarly, based on the RAF score, a level of acceptable risk score for each item is calculated by dividing the mean perceived risk score by the mean RAF score. Activities and technologies judged as currently too risky will have a lower level of acceptable risk score than the current perceived risk while those judged as safer than necessary have a level of acceptable risk score higher than the current risk.

### Task 3: Characteristics of risk

For Task 3, all participants rate each of the 18 items on two out of nine randomly assigned factors measuring different characteristics of risk. Each of the nine characteristics are measured on a seven-point scale. Participants are asked to rate all 18 items on the first scale before proceeding to the second scale. The scales are similar to that of Fischhoff et al. (1978) and Fox-Glassman and Weber (2016) and are as follows:

1. How **voluntary** is the risk? Please indicate whether you think people take on the potential risks voluntarily. If some of the risks are voluntarily undertaken and some are not, choose an appropriate rating toward the center of the scale.  
   (1 = *Voluntary*, 7 = *Involuntary*).
2. How **immediate** is the risk? To what extent is the risk of death or injury immediate—or is death or injury likely to occur at some later time?  
   (1 = *Immediate*, 7 = *Delayed*).
3. **Knowledge** about the risk: To what extent are the risks known precisely by the persons who are exposed to those risks?  
   (1 = *Known precisely*, 7 = *Not known*).
4. **Scientific knowledge** about the risk: To what extent are the risks known to science?  
   (1 = *Known precisely*, 7 = *Not known*).
5. **Control over the risk:** If you are exposed to the risk of each activity or technology listed below, to what extent can you, by personal skill or diligence, avoid death or serious injury while engaging in the activity?  
   (1 = *Not controllable*, 7 = *Controllable*).
6. **Newness of the risk**: Are the risks new and novel ones, or old and familiar ones?   
   (1 = *New*, 7 = *Old*).
7. **Chronic vs. catastrophic** risk: Is it a risk that kills people one at a time **(a** chronic risk, meaning the risk is continual over time) or a risk that kills large numbers of people at once (a catastrophic risk)?   
   (1 = *Chronic*, 7 = *Catastrophic*).
8. **Common vs. dread risk**: Is it a risk that people have learned to live with (common) and can think about reasonably calmly, or is it one that people have great dread for—on the level of a gut reaction?   
   (1 = *Common*, 7 = *Dread*).
9. **Severity of consequences**: And finally, when the risk of each activity is realized in the form of a mishap or illness, how likely is it that the consequences will be fatal?   
   (1 = *Certain NOT to be fatal*, 7 = *Certain to be fatal*).

After completing all three tasks, all participants answer a number of funneling questions, provide their demographic information and are debriefed.

[*For review: The Qualtrics survey .QSF file and an exported DOCX file are provided on the OSF folder here:* <https://osf.io/hcvmz/>*. A preview link of the Qualtrics survey is provided on:* <https://hku.au1.qualtrics.com/jfe/preview/SV_9NWEMZYhh5yWUwm?Q_CHL=preview&Q_SurveyVersionID=current>.]

## Measures and data analysis strategy

We summarized the measures and statistical tests adopted in the original study,

Fox-Glassman and Weber (2016), and the current study in Table 3.

## 

## Table 3

*Comparison of data analysis strategies: Fischhoff et al. (1978), Fox-Glassman and Weber (2016), and the current study*

| **Task/Measure** | **Fischhoff et al. (1978)/Fox-Glassman and Weber (2016)** | **Current study** |
| --- | --- | --- |
| **Task 1** - **Perceived benefit and risk rating** |  |  |
| Tasks 1a/1b  Relationship between perceived benefit and perceived risk | Linear regression (item-level geometric means) | Independent samples t-tests (participant-level) |
| Task 1c (Extension 2) | Not conducted | Correlation/Linear regression (participant-level arithmetic means) |
| **Task 2** - **Risk adjustment factor** |  |  |
| Risk acceptability judgments | Frequencies of each judgment:   * “Could be riskier” * “Are currently acceptable” * “Too risky” | Frequency table of each judgment:   * “Could be much riskier” * “Could be riskier” * “Are currently acceptable” * “Too risky” * “Much too risky” |
| Risk adjustment factor score (RAF score)   * RAF scores greater than 1.0 indicate that current risk of the activity or technology is too high and must be lowered * RAF scores of 1.0 indicate that the current risk is acceptable * RAF scores less than 1.0 mean that the risk could be higher | * Risks rated “Too risky” equal the value entered to finish the statement ‘‘must be — times safer to be acceptable’’ * Risks rated “Are currently acceptable’’ equal 1.0 * Risks rated “Could be riskier” equal multiplicative reciprocal of the value that finished the statement ‘‘would be acceptable if it were — times higher’’ | Same as original |
| Level of acceptable risk score   * represents the “ideal” level of risk * activities/technologies judged as currently too risky have a lower level of acceptable risk score than the current perceived risk * activities/technologies judged as safer than necessary have level of acceptable risk score higher than the current risk | Geometric mean perceived risk score divided by geometric mean RAF score | Arithmetic mean perceived risk score divided by arithmetic mean RAF score |
| Relationship between RAF score and perceived benefit/risk | Linear regression (item-level geometric means) | Participant-level independent samples t-tests |
| Relationship between Level of acceptability score and perceived benefit/risk | Linear regression (item-level geometric means) | Participant-level independent samples t-tests |
| **Task 3** - **Characteristics of risk** |  |  |
| Relationship between 9 characteristics and perceived benefit/risk | Multiple linear regression (item-level geometric means) | Correlation (participant-level arithmetic means) |
| Relationship between 9 characteristics and RAF score | Multiple linear regression (item-level geometric means) | Correlation (participant-level arithmetic means) |
| Relationship between 9 characteristics and Level of acceptability score | Multiple linear regression (item-level geometric means) | Correlation (participant-level arithmetic means) |
| Correlation among 9 characteristics | Correlation | Not analyzed (due to 2 out 9) |
| Factor analysis of 9 characteristics | Principal components factor analysis | Not analyzed (due to 2 out 9) |

#### 

We will focus on conducting correlational analyses to examine the relationship with perceived risks, perceived benefits, and risk characteristics, yet given our design we will not conduct analyses among the characteristics as reported by Fischhoff et al. (1978) and Fox-Glassman and Weber (2016).

## Extension 1: Pandemic related items (exploratory)

We will conduct independent samples t-tests for the perceived benefit and perceived risk of the pandemic related items separately from the list of 14 items from the original study, on an individual as well as aggregate basis. We aim to determine whether participants view the relationship between perceived risk and perceived benefit differently for these items.

## Outliers and exclusions

The current replication will focus on analyzing and reporting the results of the full sample size and will not attempt to identify outliers. We will not be making any corrections to raw data, and we will be reporting results for both pre and post exclusions, with a comparison in the supplementary. Our generalized exclusion criteria are detailed in the “Exclusion criteria” subsection of supplementary materials.

## Deviations

We provided a full list of deviations and explanations for the deviations between the original study and Fox-Glassman and Weber (2016) and the present replication in Table 3 in the supplementary materials.

## Evaluation criteria for replication findings

We aimed to compare this study with the original findings in the target article. Given the number of deviations from the original we would not be able to compare effect sizes and will instead indicate whether we found a signal in support of the hypothesized effects and whether it was in the same direction as in the original study, instead of comparing effect sizes. In particular, we will conduct independent samples t-tests for Tasks 1a/1b in order to determine whether participants rate perceived risks differently than perceived benefits. We expect these t-tests to show that the two groups do indeed rate perceived risk and perceived benefit differently. For Task 1c, we will conduct correlation and linear regression analyses, however, the design of the task is fundamentally different from the original study and will not be directly comparable. Nonetheless, we do expect the analysis to reveal a negative correlation between perceived risk and perceived benefit, consistent with other studies since Fischhoff et al. (1978) (Alhakami and Slovic, 1994; Efendić et al., 2021; Finucane et al., 2000; McDaniels et al., 1997; Slovic et al., 1987).

## Replication closeness evaluation

We provided details on the classification of this study as a replication using the criteria by LeBel et al., (2018) criteria in Table 4 below (see section “Replication evaluation” in the supplementary). We summarized this study as being between a close and a far replication.

## 

## Table 4

*Classification of the replication of Fischoff et al. (1978), based on LeBel et al. (2018)*

|  |  |  |
| --- | --- | --- |
| **Design facet** | **Replication** | **Details of deviation** |
| Effect/hypothesis | Same |  |
| IV/DV construct | Same |  |
| IV/DV operationalization | Similar | Slight adjustment to the definition of “risk” as compared to the original, and to the scales used. See Table 3 in the supplementary materials. |
| IV/DV stimuli | Similar | Changes made to list of items to reduce duration per participant and cognitive burden. See Table 3 in the supplementary materials. |
| Procedural details | Similar | Changes were made to the procedure in order to reduce duration per participant and cognitive burden. Changes were also made to data analysis strategy. See Table 3 in the supplementary materials. |
| Physical settings | Different | Original was performed on paper, whereas in the current study, participants completed a Qualtrics survey online. |
| Contextual variables | Different |  |
| Population (e.g. age) | Different | Original participants were members of Eugene Oregon League of Women Voters, whereas current study participants were online MTurk workers with a wider demographic range. |
| Replication classification | Close/Far replication |  |

## 

## 

## Table 5

*Classification of the replication of Fox-Glassman and Weber (2016), based on LeBel et al. (2018)*

|  |  |  |
| --- | --- | --- |
| **Design facet** | **Replication** | **Details of deviation** |
| Effect/hypothesis | Same |  |
| IV/DV construct | Same |  |
| IV/DV operationalization | Similar | Slight adjustment to the definition of “risk” as compared to the original, and to the scales used. See Table 3 in the supplementary materials. |
| IV/DV stimuli | Similar | Changes made to list of items to reduce duration per participant and cognitive burden. See Table 3 in the supplementary materials. |
| Procedural details | Similar | Changes were made to the procedure in order to reduce duration per participant and cognitive burden. Changes were also made to data analysis strategy. See Table 3 in the supplementary materials. |
| Physical settings | Same |  |
| Contextual variables | Same |  |
| Population (e.g. age) | Same |  |
| Replication classification | Close/Far replication |  |

# 

# Results

**[IMPORTANT:   
We provided partial results in the following sections written using a randomized dataset produced by Qualtrics to simulate what these sections will look like after data collection. These will be updated following the data collection. No pre-registration or data collection took place yet.]**

## Perceived risk and perceived benefit: Tasks 1a/1b (replication)

For participants performing Tasks 1a and 1b, we will summarize the means and other descriptive statistics for ratings of perceived risk and perceived benefit for each of the 18 items in Table 5. Kendall’s coefficient of concordance will be calculated for both perceived risk and perceived benefit and to indicate inter-participant agreement. [For purposes of the current review, we provided partial results based on dummy data generated in Qualtrics (*N* = 1000) in Table 1.]

[Some of the tables below are meant to simulate the tables that would appear after data collection. Given that the simulated dataset is random noise, we did not update the table with that information.]

## Table 5 *Mean judgments of risk and benefit from 18 activities and technologies*

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Activity/Technology | Perceived  benefit (*n* = [X]) | | Perceived risk  (*n* = [X]) | | Mean Risk adjustment factor score1 | | | | Acceptable level of risk score2 | | |  |
| (Benefit raters) (*n* = [X]) | | (Risk raters) (*n* = [X]) | | (Benefit raters) (*n* = [X]) | | (Risk raters) (*n* = [X]) | |
|  | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Alcoholic beverages |  |  |  |  |  |  |  |  |  |  |  |  |
| Contraceptives |  |  |  |  |  |  |  |  |  |  |  |  |
| Covid-19 vaccines\* |  |  |  |  |  |  |  |  |  |  |  |  |
| Electric power |  |  |  |  |  |  |  |  |  |  |  |  |
| Experimentation with biological viruses\* |  |  |  |  |  |  |  |  |  |  |  |  |
| Food preservatives |  |  |  |  |  |  |  |  |  |  |  |  |
| General aviation |  |  |  |  |  |  |  |  |  |  |  |  |
| General vaccinations\*\* |  |  |  |  |  |  |  |  |  |  |  |  |
| Handguns |  |  |  |  |  |  |  |  |  |  |  |  |
| Lockdowns\*  (Covid-19 pandemic) |  |  |  |  |  |  |  |  |  |  |  |  |
| Motor vehicles |  |  |  |  |  |  |  |  |  |  |  |  |
| Nuclear power |  |  |  |  |  |  |  |  |  |  |  |  |
| Pesticides |  |  |  |  |  |  |  |  |  |  |  |  |
| Prescription antibiotics |  |  |  |  |  |  |  |  |  |  |  |  |
| Smoking |  |  |  |  |  |  |  |  |  |  |  |  |
| Social distancing\* (Covid-19 pandemic) |  |  |  |  |  |  |  |  |  |  |  |  |
| Surgery |  |  |  |  |  |  |  |  |  |  |  |  |
| X-rays |  |  |  |  |  |  |  |  |  |  |  |  |
| All ratings |  |  |  |  |  |  |  |  |  |  |  |  |
| Coefficient of concordance (W) |  |  |  |  |  |  |  |  |  |  |  |  |

*Notes.*

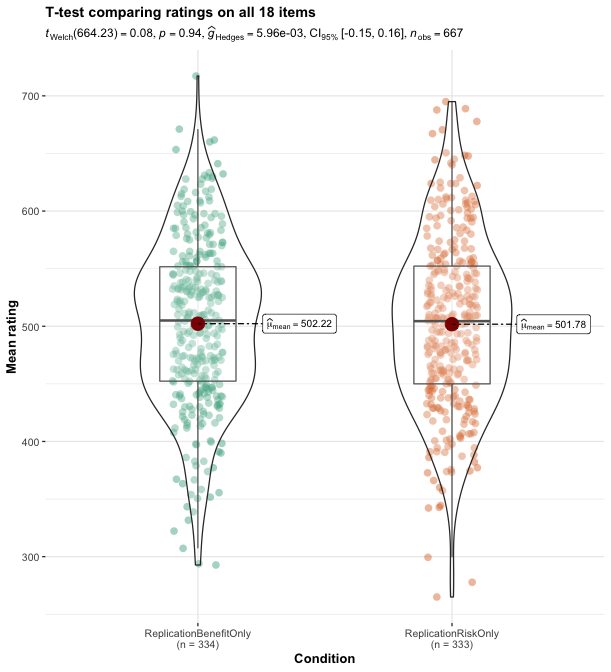
\* New Covid-19 pandemic related item.

\*\* Modified version of item from Fischhoff et al. (1978).

1. Values greater than one mean that the item should be safer; values less than one mean that the item could be riskier. See “Methods - Measures and data analysis strategy - Task 2 - Risk adjustment factor for details on calculation of this measure.
2. Acceptable levels of risk were calculated by dividing perceived risk by RAF score and represents the “ideal level of acceptable risk”.
3. SD = standard deviation

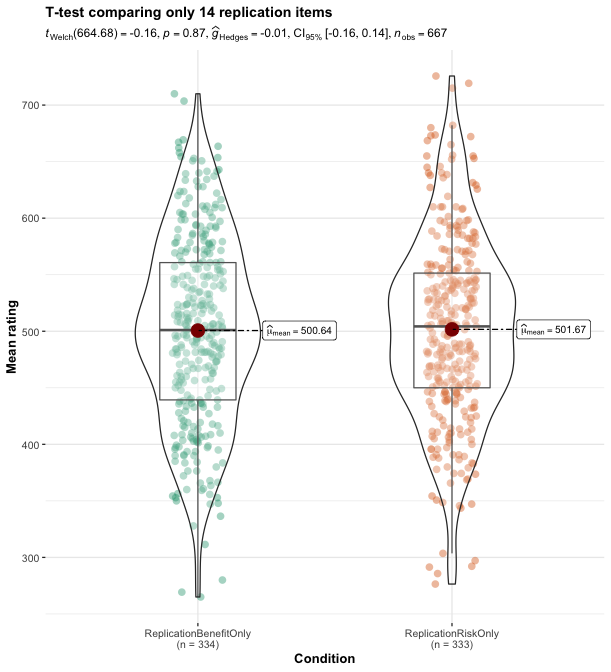
We will conduct independent samples t-tests on the mean perceived benefits and perceived risk scores for all 18 items. [For purposes of the current review, analysis of dummy data generated in Qualtrics (*N* = 1000) found no support for the negative correlation between risks and benefit.]

## Figure 1



We will conduct independent samples t-tests on the mean perceived benefits and perceived risk scores for all 14 of the items from Fischhoff et al. (1978) reproduced for this study. [For purposes of the current review, analysis of dummy data generated in Qualtrics (*N* = 1000) found no support for the negative correlation between risks and benefit.]

## Figure 2



We will also conduct independent samples t-tests on each item individually to determine the relative differences between risk and benefits ratings across the different items. We will summarize all t-test results by grouping and by individual item in Table [X].

## 

## Table 6 *Comparing perceived benefit and perceived risk: Summary of t-test analyses*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Independent samples t-test | Welch’s *t* statistic | df | *p* | Mean difference | SE difference | Cohen’s *d* | 95% CI |
| Perceived benefit and risk |  |  |  |  |  |  |  |
| All 18 items |  |  |  |  |  |  | / |
|
| 14 replication items |  |  |  |  |  |  |  |
| [Each item individually] |  |  |  |  |  |  |  |

*Note.* df indicates degree of freedom, SE indicates standard error, and CI indicates confidence interval

### Risk adjustment factor

#### Risk adjustment score

We will summarize the mean RAF scores for each of the items separately for benefit raters and risk raters in Table 1 above. [For purposes of the current review, we provided partial results based on dummy data generated in Qualtrics (*N* = 1000) in Table 1]. We will conduct independent samples t-tests consistent with the previous analysis on perceived risk and benefit, comparing the 18 individual items, 18 aggregated items, 14 original items aggregated by the Risk rater and Benefit rater groups.

#### Level of acceptable risk score

We will summarize the mean level of acceptable risk scores for each of the items separately for benefit raters and risk raters in Table [5] above. [For purposes of the current review, we provided partial results based on dummy data generated in Qualtrics (*N* = 1000) in Table 1]. We will conduct independent samples t-tests consistent with the previous analysis on perceived risk and benefit, comparing the 18 individual items, 18 aggregated items, 14 original items aggregated by the Risk rater and Benefit rater groups.

### Characteristics of risk

We will summarize the summary statistics for ratings of the 9 characteristics of risk for each of the 18 activities and technologies in Table [8] below. We will calculate Kendall’s coefficient of concordance for all scales to indicate inter-participant agreement.

## 

## Table 7 *Characteristics of risk summary statistics table*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Activity/Technology | Voluntariness  1 = voluntary  (*n* = [X]) | | Immediacy  1 = immediate  (*n* = [X]) | | Known to exposed  1 = known precisely  (*n* = [X]) | | Known to science  1 = known precisely  (*n* = [X]) | | Controllability  1 = uncontrollable  (*n* = [X]) | | Newness  1 = new  (*n* = [X]) | | Chronic/Catastrophic  1 = chronic  (*n* = [X]) | | Common/Dread  1 = common  (*n* = [X]) | | Severity  1 = certain to be fatal  (*n* = [X]) | |
|
|  | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Alcoholic beverages |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Contraceptives |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Covid-19 vaccines\* |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Electric power |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Experimentation with biological viruses\* |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Food preservatives |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| General aviation |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| General vaccinations\*\* |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Handguns |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Lockdowns\*  (Covid-19 pandemic) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Motor vehicles |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Nuclear power |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Pesticides |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Prescription antibiotics |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Smoking |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Social distancing\* (Covid-19 pandemic) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Surgery |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| X-rays |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| All ratings |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Coefficient of concordance (W) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

*Notes.*

\* New Covid-19 pandemic related item.

\*\* Modified version of item from Fischhoff et al. (1978).

1. SD = standard deviation.

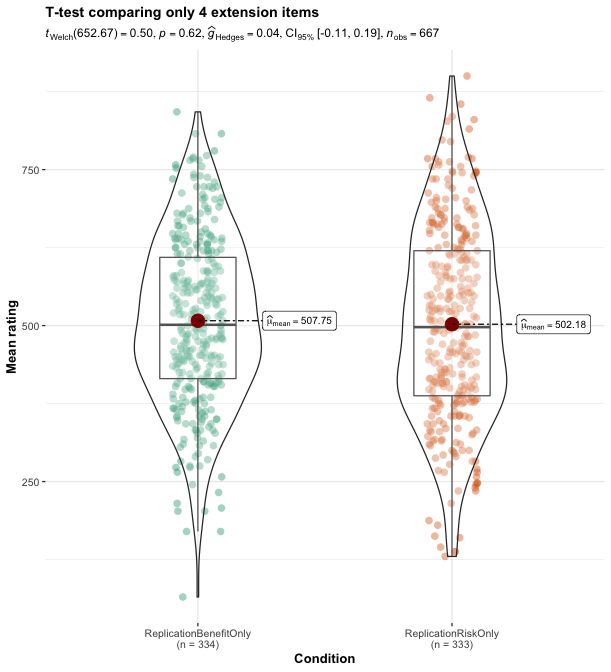
### For analysis of characteristics of risk, we will prepare correlations matrices showing the relationships between perceived risk and perceived benefit and the 9 characteristics, between RAF scores and the 9 characteristics, and the level of acceptable risk scores and the 9 characteristics.

### Extension 1: Pandemic related items

#### Perceived risk and perceived benefit

For purposes of our extension, we will conduct independent samples t-tests on the 4 pandemic related items separately from the other 14 items, both in the aggregate and individually. [For purposes of the current review, analysis of dummy data generated in Qualtrics (*N* = 1000) found no support for the negative correlation between risks and benefit for these extension items.]

Figure [X]



[We will summarize all t-test results for the pandemic related items by grouping and by individual item in Table [8].]

## 

## Table 8

## *Summary of all t-tests results for perceived benefit and perceived risk scores - Covid-19 Items*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Independent samples t-test | Welch’s *t* statistic | df | *p* | Mean difference | SE difference | Cohen’s *d* | 95% CI |
| Perceived benefit and risk |  |  |  |  |  |  |  |
| Pandemic related items |  |  |  |  |  |  | / |
|
| [Each item individually] |  |  |  |  |  |  |  |

#### Risk adjustment factor

## We will conduct analyses on the RAF scores and Level of acceptable risk scores separately for the pandemic related items consistent with the previous analysis.

## Perceived risk and perceived benefit: Extension 2 (Task 1c)

We will summarize summary statistics for participants performing Task 1c in a table similar to Table [5] above. We will then test for support for the negative correlation between risk and benefit by conducting correlation and linear regression on the participant-level perceived risk and perceived benefit ratings for each item.

### Characteristics of risk

We will also conduct analyses consistent with the previous characteristics of risk section for the participants that completed Task 1c.

## Comparing this study to original findings

Since the simulated dataset generated random noise, the comparison between this study and findings in the original is irrelevant, and will only be completed after data collection. We will aim to compare the results of the replication to the original findings to the extent possible based on the criteria by LeBel et al. (2019) (see supplementary materials for more details). However, given the number of deviations from the original we will not be able to compare effect sizes and will instead indicate whether we found a signal in support of the hypothesized effects and whether it was in the same direction as in the original study, instead of comparing effect sizes. In particular, we will conduct independent samples t-tests for Tasks 1a/1b in order to determine whether participants rate perceived risks differently than perceived benefits. For Task 1c, we will conduct correlation and linear regression analyses, however, the design of the task is fundamentally different from the original study and will not be directly comparable.

# Discussion

*[Please note that the discussion section at this stage is only meant as a teaser placeholder of identified topics to be discussed and completed in Stage 2 following data collection]*

## Limitations

### Comparison across studies

We made many changes to the target article’s and Fox-Glassman and Weber (2016)’s study design. These departures limited our ability to compare between the current study and those two studies. Our list of items was primarily based on the same items used in Fischhoff et al. (1978) and Fox-Glassman and Weber (2016), yet it is possible that in the intervening years since these studies, people’s understanding of these items and their attitudes toward their risks and benefits have changed. Moreover, reporting of risk preferences may be sensitive to context, choice options, and elicitation methods (Frey et al., 2017; Jusev et al., 2020). We therefore advise caution regarding drawing any strong conclusions regarding comparisons of our results and these two studies.

### Order effects and confounds

We did not randomize the presentation of survey items to participants, instead grouping items with other similar items to reduce cognitive burden across tasks. It is possible that this may have produced ordering effects. In addition, participants in Task 1a (the risk raters) and Task 1b (the benefit raters) completed Tasks 2 and 3 whereas participants in Task 1c (rating both risks and benefits) only completed Task 3 without completing Task 2. It is possible that completing Task 2 somehow affects how participants respond to Task 3, and this can be addressed with an exploratory analysis comparing Task 3 completed by Task 1c participants to Task 3 completed by Tasks 1a and 1b participants.

### COVID-19 Generalizability

Our study was conducted in May of 2022, while the COVID-19 pandemic was still very much ongoing. During this time, lockdowns, quarantines, mask-wearing policies, not to mention the medical impact of the pandemic, were impacting the general public. Accordingly, the results of this study pertaining to the COVID-19 pandemic may be constrained in terms of its generalizability.

## Future directions

### Differences between Perceived Risk and Perceived Benefit Ratings

### The use of independent sample t-tests to analyze results from Tasks 1a/1b and Task 2 was primarily exploratory. Results indicated that participants [to come]. These results imply that [to come]. Future research regarding the perceived risk and perceived benefit relationship may consider these findings and their potential further implications.

### [Possibly discuss the promising future direction of comparing the between-subject and within-subject designs in the unified data collection.]

### Numeracy and risk judgment

### Recent research has indicated that people have difficulty with numerical expression of their own risk judgments. In particular, Raude et al. (2021) found that individual numeracy plays an important role in the magnitude by which people overestimate the perceived riskiness of certain common illnesses. In the current study, we did not measure participants' individual numeracy and as a result, we are unable to report if the effects we observed are influenced by numeracy. Future research adopting a similar methodology may consider an individual numeracy measure to test whether, as has been shown elsewhere, numeracy affects the perceived risk and perceived benefit relationship.

# References

Alhakami, A. S., & Slovic, P. (1994). A Psychological Study of the Inverse Relationship Between Perceived Risk and Perceived Benefit. *Risk Analysis, 14(6), 1085–1096.* doi:10.1111/j.1539-6924.1994.tb00080.x

Brown, R., Coventry, L. & Pepper, G. COVID-19: the relationship between perceptions of risk and behaviours during lockdown. *Journal of Public Health (Berl.)* (2021). <https://doi.org/10.1007/s10389-021-01543-9>

Efendić, E., Chandrashekar, S. P., Lee, C. S., Yeung, L. Y., Kim, M. J., Lee, C. Y., & Feldman, G. (2021). Risky Therefore Not Beneficial: Replication and Extension of Finucane et al.’s (2000) Affect Heuristic Experiment. *Social Psychological and Personality Science.* <https://doi.org/10.1177/19485506211056761>

Finucane, M.L., Alhakami, A.S., Slovic, P., & Johnson, S.M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making, 13*, 1-17.<https://doi.org/10.1002/(SICI)1099-0771(200001/03)13:1>[<1::AID-BDM333>3.0.CO;2-S](https://psycnet.apa.org/doi/10.1002/(SICI)1099-0771(200001/03)13:1%3C1::AID-BDM333%3E3.0.CO;2-S)

Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., & Combs, B. (1978). *How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. Policy Sciences, 9(2), 127–152.* [*https://doi.org/10.1007/BF00143739*](https://doi.org/10.1007/BF00143739)

Fox-Glassman, K. T., & Weber, E. U. (2016). What makes risk acceptable? Revisiting the 1978 psychological dimensions of perceptions of technological risks. *Journal of Mathematical Psychology, 75,* 157 169.<https://doi.org/10.1016/j.jmp.2016.05.003>

Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science advances*, *3*(10), e1701381. <https://doi.org/10.1126/sciadv.1701381>

Gigerenzer, G. (2004). Dread Risk, September 11, and Fatal Traffic Accidents. *Psychological Science*, *15*(4), 286–287. <https://doi.org/10.1111/j.0956-7976.2004.00668.x>

Heirene, R. (2020, January 7). A call for replications of addiction research: Which studies should we replicate & what constitutes a “successful” replication?. <https://doi.org/10.31234/osf.io/xzmn4>

Kasperson, R. E., Renn, O., Slovic, P., Brown, H. S., Emel, J., Goble, R., … Ratick, S. (1988). *The Social Amplification of Risk: A Conceptual Framework. Risk Analysis, 8(2), 177–187.* <https://doi.org/10.1111/j.1539-6924.1988.tb01168.x>

Koonce, L., McAnally, M. L., & Mercer, M. (2005). How do investors judge the risk of financial items? The Accounting Review, 80(1), 221–241. <http://dx.doi.org/10.2308/accr.2005.80.1.221>

Kusev, P., van Schaik, P., Martin, R., Hall, L., and Johansson, P. (2020). Preference reversals during risk elicitation. *Journal of Experimental Psychology: General* 149, 585–589. <http://dx.doi.org/10.1037/xge0000655>

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior research methods*, 49(2), 433-442. <https://doi.org/10.3758/s13428-016-0727-z>

LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, *1*, 389-402. [https://doi.org/10.1177/2515245918787489](https://psycnet.apa.org/doi/10.1177/2515245918787489)

LeBel, E. P., Vanpaemel, W., Cheung, I., & Campbell, L. (2019). A brief guide to evaluate replications. *Meta-Psychology*, 3, 1-9. <https://doi.org/10.15626/MP.2018.843>

Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. Revue Internationale de Psychologie Sociale, 32(1). <https://doi.org/10.5334/irsp.289>

McDaniels, T. L., Axelrod, L. J., Cavanagh, N. S., & Slovic, P. (1997). Perception of ecological risk to water environments. *Risk Analysis, 17*(3), 341–352.<https://doi.org/10.1111/j.1539-6924.1997.tb00872.x>

Nosek, B. A., & Errington, T. M. (2020). What is replication?. *PLOS Biology*, 18(3), e3000691. <https://doi.org/10.1371/journal.pbio.3000691>

Raude, J., Xiao, C., & Crépey, P. (2021, April 10). Revisiting the primary bias: the role of innumeracy in the misperception of prevalence of common illnesses. *PsyArXiv*, 10 Apr. 2021. <https://doi.org/10.31234/osf.io/fnm5g>

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. The Journal of Finance, 19(3), 425. <https://doi.org/10.2307/2977928>

Skagerlund, K., Forsblad, M., Slovic, P., & Västfjäll, D. (2020). *The Affect Heuristic and Risk Perception – Stability Across Elicitation Methods and Individual Cognitive Abilities. Frontiers in Psychology, 11.* doi:10.3389/fpsyg.2020.00970

Slovic, P. (1986). Informing and Educating the Public About Risk. *Risk Analysis, 6(4), 403–415.* <https://doi.org/10.1111/j.1539-6924.1986.tb00953.x>

Slovic, P. (1987). Perception of risk. *Science*, 236(4799), 280–285.<https://doi.org/10.1126/science.3563507>

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). *The affect heuristic. European Journal of Operational Research, 177(3), 1333–1352.* <https://doi.org/10.1016/j.ejor.2005.04.006>