Answer to recommender and referees

## Removing barriers to plant-based diets: assisting doctors with vegan patients

*Romain Espinosa, Thibaut Arpinon, Paco Maginot, Sébastien Demange, and Florimond Peureux*

*November 2021*

First of all, we would like to thank the recommender and the referees for their detailed reading and the thoughtful comments on the previous version of our work. We were happy to read that the recommender and the referees think our paper addresses an interesting topic. We did our best to take their comment into account when revising our pre-registered report. Most, if not all, comments were justified and we believe that the report improved during the process.

In the following pages, the italic paragraphs are copy-pasted from the reports, our answers are displayed in normal font.

***Please note that the version that track the changes made on the manuscript is available here:***

[*https://osf.io/9qeny/?view_only=66eab29c7acb4aebbcec4631cbcb9217*](https://osf.io/9qeny/?view_only=66eab29c7acb4aebbcec4631cbcb9217)

## **General comments**.

Several questions were raised in the reports regarding the validity of our instruments and tests. To address some of those concerns, we decided to run a pilot session with medical students (who almost completed their studies to become doctors, i.e., who already spent between 6 and 8 years at the University). The board of students of the University of Nantes forwarded our invitation to participate in an online survey to all students enrolled for the 7th to 9th year of study. In total, 28 students completed our survey. We briefly present in the paper the data from this pilot session. In the new version of the manuscript, we have used those data to test the validity of our dependent variables and to calibrate our power analysis. We emphasize that this sample of students is very likely to be more vegan-friendly than the average doctor in exercise as they are younger than the doctor population. Although we did not collect individual data like gender, we also expect the participants in the pilot survey to be more female. Altogether, we expect young and female doctors to be more supportive of plant-based diets than the general population of doctors. These pilot data are thus likely to overestimate the level of support for vegan diets compared to the population of doctors.

A major change in our RR concerns the statistical tests. In the original version, we proposed one-sided Wilcoxon rank-sum tests to determine whether the information campaign was successful in improving the views and the practice of doctors with respect to vegan diets. We decided to change and to use a Tobit model to test our null hypotheses for several reasons. (We also decided to stick to a frequentist approach as we have more robust expertise with these methods.)

There are two main concerns that the Wilcoxon rank-sum test was not able to address. First, it appeared (from the paper that the recommender suggested us to read) that we should not only be interested in whether the campaign has a positive impact but whether the impact of the campaign is larger than a smallest effect size of interest (SESOI). However, the effect size in a Wilcoxon rank-sum test corresponds to a rank biserial correlation, which cannot be directly compared with a SESOI on the absolute scales. So, if we are interested in comparing the estimated impact of the campaign with an effect size on the scale of the outcome variables, the Wilcoxon rank-sum test is a poor choice.

Second, the recommender and one reviewer were concerned by the fact that we did not consider the information conveyed by the data in case we were not able to reject Ho. We agree with their point and now propose to proceed in two steps. The summary of our decision process is described in the Figure below. We also added the following paragraph in the main body of the paper:

> *Hypothesis testing for the VDI and PMPI will be performed as presented in Figure 2.A (for which we expect a decrease in the scores). Hypothesis testing for the VPI is reverted and will be performed as presented in Figure 2.B (for which we expect the campaign to increase the score), and the revert will hold for VDI and PMPI (for which we expect a decrease in the scores).. Using the estimates of the Tobit (or OLS) estimation, we will perform a unilateral test to decide whether we can reject the null hypothesis $H\_{00}$ (i.e., the effect has a zero effect or an effect contrary to what is expected). Second, we will proceed to an additional unilateral test to decide whether we can reject $H\_{10}$ (i.e., the campaign has an impact larger than the SESOI). Figure 2.C summarizes the possible outcomes.*

> *If we are able to reject $H\_{00}$, we will consider that the campaign significantly increases the VPI ($\beta>0$) or significantly decreases the VDI and PMPI ($\beta<0$), and is therefore successful. However, if $H\_{10}$ is also rejected, this implies that the impact on the VPI is lower than the SESOI ($\beta<\pi$) or greater than the SESOI ($\beta>\pi$) for the VDI and PMPI, which makes it only weakly successful. If $H\_{00}$ is not rejected, we cannot say that the campaign was successful. In case $H\_{10}$ is also rejected, we know that the campaign has an effect on the VPI lower than the SESOI or greater than the SESOI for the VDI and PMPI, which can be even zero: we thus conclude that the campaign was unsuccessful. On the contrary, if $H\_{10}$ is not rejected, we have inconclusive evidence as the campaign can be both successful and unsuccessful. Figure 2.C summarizes the possible outcomes.*

**A)**

**VDI and PMPI**

| Test | Decision Rule |
|------|---------------|
| $H_{00} : \beta \geq 0$ | If $\dfrac{\hat{\beta}}{\widehat{SE}_{\hat{\beta}}} < t_{1-\frac{\alpha}{3}}$, $H_{00}$ rejected |
| $H_{01} : \beta < 0$ | If $\dfrac{\hat{\beta}}{\widehat{SE}_{\hat{\beta}}} \geq t_{1-\frac{\alpha}{3}}$, $H_{00}$ not rejected |
| $H_{10} : \beta \leq \pi$ | If $\dfrac{\hat{\beta}-\pi}{\widehat{SE}_{\hat{\beta}}} > -t_{\frac{\alpha}{3}}$, $H_{10}$ rejected |
| $H_{11} : \beta > \pi$ | If $\dfrac{\hat{\beta}-\pi}{\widehat{SE}_{\hat{\beta}}} \leq -t_{\frac{\alpha}{3}}$, $H_{10}$ not rejected |

**B)**

**VPI**

| Test | Decision Rule |
|------|---------------|
| $H_{00} : \beta \leq 0$ | If $\dfrac{\hat{\beta}}{\widehat{SE}_{\hat{\beta}}} > t_{1-\frac{\alpha}{3}}$, $H_{00}$ rejected |
| $H_{01} : \beta > 0$ | If $\dfrac{\hat{\beta}}{\widehat{SE}_{\hat{\beta}}} \leq t_{1-\frac{\alpha}{3}}$, $H_{00}$ not rejected |
| $H_{10} : \beta \geq \pi$ | If $\dfrac{\hat{\beta}-\pi}{\widehat{SE}_{\hat{\beta}}} < -t_{\frac{\alpha}{3}}$, $H_{10}$ rejected |
| $H_{11} : \beta < \pi$ | If $\dfrac{\hat{\beta}-\pi}{\widehat{SE}_{\hat{\beta}}} \geq -t_{\frac{\alpha}{3}}$, $H_{10}$ not rejected |

**C)**

**Outcome**

| | $H_{00}$ not rejected | $H_{00}$ rejected |
|---|---|---|
| $H_{10}$ not rejected | Inconclusive | Successful campaign |
| $H_{10}$ rejected | Unsuccessful campaign | Weakly successful campaign |

**Figure 2** – Hypothesis testing for the main outcome variables and interpretation of possible results. **A)** Unilateral hypothesis testing and decision rule for the *Veganism Disapproval Index* and *Proper Medical Practice Index* using the regression coefficient ß and smallest effect size of interest π. **B)** Unilateral hypothesis testing and decision rule for the *Veganism Promotion Index* using the regression coefficient ß and smallest effect size of interest π. **C)** All possible outcomes of hypothesis testing and their interpretation in terms of information campaign impact.

Last, we explain in the RR that the Tobit model will be used only if we have observations at the upper / lower limits of the scale. If it is not the case, we commit to use OLS to estimate the treatment impact.

**Recommender:** Zoltan Dienes

*One question, raised by Aldoh, is whether you would want to pilot your scales to establish their reliability or validity. Or else, as suggested by Palfi, introduce into the main study "outcome neutral tests" - or estimates - of their scale properties, so that main conclusions are conditional on these tests showing adequate psychometric properties. A minor point: For the first scale, how about asking the*

We agree with the suggestion of referee Palfi. As we explained above, we collected pilot data to test the reliability of the scales. Note that, before piloting our design, we changed the scales as recommended, using percentages from 0% to 100% by 10 percentage-point increments.

- Regarding the VDI scale, the Cronbach's alpha is equal to 0.80, which is satisfactory.
- Regarding the ceiling and floor effect:
    - We have no observation at the bounds for VDI.
    - We have no observation at the bounds for the PMPI.
    - We have a large share of participants who give all the money to the charity (61%) and a negligible share who gives nothing (7%). (VPI)

Given these figures, we believe that ceiling / floor effects are not an issue for VDI and PMPI. However, the ceiling effects can be an issue for VPI.

Overall, we take into account the potential ceiling/floor effects as follows. First, we decided to use a Tobit model to estimate the treatment effect, which accounts for the inflation at lower and upper limits of the scales. Second, we introduced outcome neutral tests to ensure that we are able to detect an effect in case there is a large amount of ceiling/floor data. We wrote the following in the manuscript:

> The effect of the information campaign on the three outcome variables will be estimated using a Tobit model to take into account the possible inflation of observations at the lower and upper limits of the scales. In case there are no such observations, the treatment effect will be estimated using OLS. We use a significance threshold of 0.05 but we apply a Bonferroni correction for multiple hypothesis testing and retain alpha=0.0167 (three outcome variables).

> We set up outcome neutral tests to anticipate potential ceiling or floor effects. Prior to hypothesis testing, we will run the above power analysis again using data from the control group. We commit to analyze only the outcome variables for which we are able to estimate a medium effect (Cohen's d of 0.5).

Thank you for pointing this out. Following your paper on "Obtaining evidence for no effect", we decided to change our perspective a bit. We now define the smallest effect size of interest

(SESOI=π), and test two null hypotheses (please see the "General Comments" section of this letter). To define the SESOI, we consider a Cohen's d smaller than 0.5. In other words, we will consider that the campaign is successful if it has an effect larger than a Cohen's d of 0.5. The choice of 0.5 as a cutoff is motivated by the fact that there is no good economic reasoning that could strongly argue for a specific cutoff. We thus choose this statistical cutoff as it is standard in the literature for a "medium" effect, even though we noted your criticisms in your paper about evidence for no effect.

> *In terms of power, you have taken as a starting point your sample size, then asked what effect size that implies for an 80% power. As both Palfi and Aldoh asked, why should we be concerned specifically about an effect of 0.1? This point is important in terms of whether a non-significant result would refute your hypothesis. Would a non-sig result allow the conclusion you state follows from it: "the information campaign fails to improving doctors' views of plant-based diets." Only if power was calculated with respect to a minimally interesting effect for your research problem. Aldoh also points out that conclusions based only on the logic of power do not take into account the data as they are actually observed; for example, a significant result may be less than 0.1; and a non-significant result may come with a confidence interval that extends beyond 0.1.*
>
> *It is up to you what inferential approach you wish to adopt (i..e you can stick with a NP power approach), but some comment on this would be helpful. In effect Aldoh is raising the possibility of an equivalence region approach. It would still require justifying a minimally interesting effect size. Palfi wonders if Bayes factors may be helpful in this regard. Then one needs to say not what is the minimally interesting effect, but what effect is predicted by a theory. (The theory could be e.g. that any difference is possible within the range of the scale, though smaller effects are more likely than bigger ones.) (Some ideas here may help for any of these approaches: https://psyarxiv.com/yc7s5/ )*
>
> *If justifying a minimally interesting effect or a predicted effect seems difficult, that may be because this is a situation where one should just estimate the effect size with its 95% CI. Approached in this way, your conclusion would not be that the intervention does or does not work; but that the estimate of how well it works is such and such.*
>
> *The exact resolution of this inferential issue could go in several directions. Both myself and the reviewers thought this needed more work however.*

Thank you again for this comment. As we detailed above, we significantly changed our approach in the way you and the reviewers proposed. We hope that the changes respond to your concerns.

> *Tables S1 and S2:Column labeled "Beta" - you mean "power"?*

Yes, thank you.

**Report #1:** Joshua Tasoff

---

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

*Valid, very interesting and very important.*

*1B. The logic, rationale, and plausibility of the proposed hypotheses (where a submission proposes hypotheses)*

*The project is plausible, logical, and presented clearly.*

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

*The approach is sound and feasible.*

*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*

*Everything necessary for replication is available.*

*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).*

*I think the authors have considered reasonable scenarios. Unfortunately, a messy world produces messy data. There may be unforeseen factors that may get in the way of logistics or make the desired tests less ideal than they appear now. But I believe the authors have done their due diligence.*

---

Thank you very much for this positive feedback!

**Report #2:** Bence Palfi

---

*The manuscript aims to test whether providing information about the health benefits of plant-based diets to physicians can change their attitudes and behaviour towards veganism. I believe that the proposed research question is valid and very promising, it would be certainly intriguing to explore the potential of*

Thank you very much for pointing this out. We omitted the information about the recruitment process, which is indeed crucial for the external validity of our studies and the generalisability of our results. We added the following paragraph:

> "The participants are selected from the polling institute's database of doctor participants and are not recruited for this specific experiment. They are randomly allocated in one of the two conditions (baseline vs. treatment), and receive a fixed amount of money (€18) determined by the standard practice of the polling institute."

> *Will the participants be randomly assigned to the conditions? This is critical to the comparability of the conditions.*

Yes, they are. Please, see our answer to the previous comment.

> *There are some redundancies: the size of the sample is mentioned multiple times as well as the information about the variables and information about the control/experimental group is scattered around the method section.*

Thank you for this remark. We did our best to eliminate the duplicates.

The variables proposed in our survey were specifically designed for this experiment. Two members of the team are doctors working on plant-based nutrition and regularly intervene to train other doctors on plant-based diets. Paco Maginot recently intervened at universities to train students who are about to finish their medical studies to become doctors. During his lecture, he asked the students whether they would prescribe some biological exams from a list of exams. The list that we report in the manuscript is largely inspired from the list of biological exams he used in his class survey. (This is how we calibrated our power analysis in the previous version of the manuscript.) We decided to add T9 and T10 in the list because our two doctors reported that many of their colleagues would be willing to prescribe them, although this would not be considered as good or bad practice.

Regarding T9, some doctors would be willing to test for the concentration of B12 through a blood exam. Checking for B12 is a good practice for vegan patients as there is no B12 in plant-based foods and patients must therefore take supplements for it. However, a blood test is not informative if the patient eats seaweed. So, while checking for B12 is a good practice per se, it is not ideal to do it through a blood test (the urine test in T1 is more robust). We thus decided to leave it out of the PMPI but still to include it as several doctors might ask for it.

Regarding T10, it is a very standard test that many doctors prescribe in various situations. The two doctors in our team believe that their colleagues might ask for this blood test whenever they ask for blood exams. Prescribing T10 would thus reflect a general practice that is not specific to vegan patients. Here again, it is not informative for PMPI but we included it in our list as it is regularly asked by doctors.

Thank you for pointing this out. We corrected it in the manuscript.

Thank you for this remark. We added the information in the main text.

> *"Based on previous data, we assume that the average probability of a positive event in the control group is 0.495". It is unclear what previous data the authors refer to. Are they referring to some pilot data or to previous studies in the field?*

In the previous version of the manuscript, we used some data to calibrate the power analysis. More specifically, we used some data that Paco Maginot, one of the co-authors, collected while he trained medical students at the university. We used the probability of prescribing each test in his sample as the probability of prescription in our analysis. Given that we have now collected pilot data, we use these pilot data to calibrate our power analysis.

> *o   The minimally interesting effect sizes (e.g., 0.1 on VDI) should be justified or put in context by comparing them to previous studies.*

Following your comments and the comments of the recommender, we changed the perspective of the test and now consider a smallest effect size of interest (SESOI). As there is no direct study on which we can rely, we consider a Cohen's d of 0.5 as the SESOI.

> ***Outcome neutral tests*** *play an important role in RRs to ensure the quality of the data. I think all three central tests could have a corresponding outcome neutral test to ensure that there is no ceiling or floor effect in the control groups. You could run a two-sided Wilcoxon test in the control group for all three variables.*

Thank you for pointing this out. Following your comment we proceeded to several changes. First, we ran a pilot session to test the robustness of our outcome variables and elicit the presence of floor / ceiling effects. In our pilot sample:

- o   We have no observation at the bounds for VDI.
- o   We have no observation at the bounds for the PMPI.
- o    We have a large share of participants who give all the money to the charity (61%) and a negligible share who gives nothing (7%). (VPI)

Following these results, we decided to opt for a Tobit model, as it can take into account the inflation at the upper and lower limits of the scales.

Last, we decided to introduce outcome neutral tests as you suggested. We added the following paragraph:

> *We set up outcome neutral tests to anticipate potential ceiling or floor effects. Prior to hypothesis testing, we will run the above power analysis again using data from the control group. We commit to analyze only the outcome variables for which we are able to estimate a medium effect (Cohen's d of 0.5).*

> *The first central test investigating the extent of attitude change has a potentially important follow-up test that, in my opinion, could also be preregistered. Elevating the attitudes towards plant-based diets is useful however, the intervention may not be good enough if the attitudes remain negative in the experimental group (VDI smaller than 0.5). This is of course given that the attitudes were negative in the control group and the intervention managed to elevate them.*

Thank you for this suggestion. We believe that an information campaign can still be useful even if it is not sufficient to make the average views about plant-based diet positive. It might indeed take several expositions to information campaigns, or a combination with other tools (like new dietary guidelines) to achieve this. Our focus is here to estimate the impact of a single exposure to this type of information campaign. So, we prefer not to expand the set of pre-registered hypotheses.

> *The authors may find it difficult to interpret some of their results given that they are non-significant. I recommend the inclusion of Bayesian analyses (the Bayes factor) so that the authors can distinguish between inconclusive results and clear evidence for the null. Bayes factors can be included conditionally (in case a test is non-significant) or they can be run for every single statistical test. JASP (https://jasp-stats.org/) offers a simple way to run Bayesian Mann-Whitney and Wilcoxon tests.*

Thank you for this comment. Following your advice and the editor's remarks, we decided to change our perspective such as to include the information that can support a null effect. Thank you also for pointing out this resource, which we did not know. After looking at the alternatives, we prefer to stick to a frequentist approach as we have more expertise in this area and the recommender left this window open.

We summarized the new approach in the *General Comments* at the beginning of the letter.

*In general, I've found the introduction to be very informative, concise, and well-argued. However, it would be great to see some paragraphs on the psychology of attitude/behaviour change via information campaigns.*

Thank you for this suggestion. We have added the following paragraph in the introduction.

> "Information campaigns can be effective tools to educate, update beliefs and induce behavioural changes, may it be in politics,[38] environmental behaviors[39] or health.[40,41] Specifically, health information campaigns have been proven successful to educate the general population on topics such as nutrition[42], HIV,[43] strokes,[44] vaccines,[45] and more recently guidelines during the COVID-19 pandemic.[46] Additionally, informing professionals about the latest scientific advances and methods, in particular doctors, can lead to better practices and updated knowledge.[47] "

*I think that the end of the introduction does not need to mention the specific statistical analyses that will be used to test the hypotheses. It is enough to specify the research question and the hypotheses. The statistical analyses are described in detail later.*

Thank you for this remark. We proceeded as suggested.

*"Physician" may be a more precise term than "doctor". More importantly, am I right in thinking that the participants will be primary care physicians (general practitioners) and not any kind of doctors? I think this should be clarified.*

Thanks for your comment. We originally wrote "physicians" but our proofreader advised us to write "doctors" instead. (As he comes from England, there might be some cross-country variations that we are not aware of.) Following your remark, we decided to write in the abstract and introduction that our study concerns general practitioners.

*In the description of Figure 1, the authors claim that the p values in a previous study were significant at the 1% level (in fact, $p < 0.001$ is also significant at the 0.1% level), but they use the 5% level in their own*

*study. Alpha levels should be specified before data collection, and I suspect that the authors of the mentioned study did not intend to use the 1% threshold, so it is better to stick with the traditional threshold of 0.05 when describing the study.*

We changed the p-values in our graph from "p<0.001" to "p<0.05".

**Report #3:** Alaa Aldoh

*Overview*

*Doctors' possible misperceptions of plant-based diets may compromise their relationships with patients, and the willingness of newly vegetarian/vegan patients to continue eating plant-based foods. The article suggests that doctors in France are not always willing to learn about plant-based nutrition as a result of barriers including cognitive factors and otherwise. The article outlines plans for a randomised controlled trial (N = 400) where participating doctors are assigned to one of two conditions: a) information campaign, or b) no information control. The authors hypothesize that the information campaign will: positively influence doctors' opinions of plant-based diets, increase likelihood of prescribing the correct medical test, and increase recommendations to follow plant-based diets.*

*I am not particularly aware of research in this area, but all things considered, I think this study is valuable and could have real consequences on the use of information campaigns to facilitate doctors' knowledge and promotion of plant-based diets. Some methodological decisions require further justification or elaboration to ensure the collected data adequately answers your research questions.*

*Re: Introduction*

*The introduction provides an excellent overview of doctors' perspectives on plant-based diets in France, though certain parts can be strengthened. I could not access a few of the references provided to justify the research questions, so I cannot easily evaluate this myself. I think it would be helpful to include links to theses available on the web especially when DOIs or other identifiers are not available. It is my understanding that there's been no research so far exploring the effect of information campaigns used to improve doctors' knowledge of plant-based diets, so I can understand if it not possible to use existing evidence to support the proposed study. There's good discussion of how doctors may deter patients from*

Thank you for this comment and for the references. We have added in the main text the two aforementioned papers (Cramer et al., 2017; McIntosh et al., 1995) and two other articles that show how doctors might positively influence dietary changes among their patients. Following your request, we have added the statistical test (two proportions Z- test) and the sample size (N = 1090) under the graph. We also modified the p-values to 5%. Most of the references that lack DOI are medical theses, which are not published in peer-review journals but that were accepted by juries (mandatory to become a doctor).

Thank you for these comments.

1) We agree with you on the fact that there are several scales that have been developed to assess the attitudes towards vegetarians and plant-based diets, and Corrin and Papadopoulos (2017) provide an extensive review. However, our study is more particular as we focus here on the doctor-patient relationship, which is an asymmetrical situation where one expert informs a non-expert about health risks that the patient cannot assess him/herself. We thus decided to develop our own scale.

2) Your comment and the comments of the other referee convinced us to run a pilot session to investigate the robustness of our metrics. The internal reliability of our VDI scale is relatively good as the alpha of Cronbach is equal to 0.80.

3) Following your suggestion, we changed the wording for the information campaign. We now talk about "an information campaign aimed to promote vegan diets to the general population". We add that "This campaign, developed by the CNRS and doctors, would emphasize the importance of a well-balanced vegan diet and the expected health benefits from such diet." We hope that this new formulation is clearer.

4) Regarding the active dimension of the charity giving game, we decided to slightly change the perspective following your suggestion. We contacted the polling institute and asked them whether we could incentivize the doctors. The doctors now receive €2 in addition to their fixed fee for their participation. They can keep the money for themselves or give part (or all) of it to develop the campaign. This new setting thus gives a more active role to the doctors as the money is theirs.

5) Last, we made the question about the consumption levels more explicit. It now reads:

   *What is your opinion regarding the level of consumption by the French population of the following foods? (From Strongly insufficient (i.e., they eat too little of this food) to Strongly excessive (i.e., they eat too much of this food.)*

---

*Presentation of measures*

*Regarding the scales presented on screen #4, I think it might be confusing to vary the valence of the anchors across questions. I think it would be better to keep the negative anchors on the same side for each question. The booklet used to provide information relates to both vegetarian and vegan diets, but the measures used refer to vegan diets specifically. I think it would be better either to adjust the prompts, or to clarify in your discussion that this research applies to vegan diets specifically. Perhaps discuss this limitation in your report, as doctors may have more favourable attitudes towards vegetarian diets in comparison to vegan ones.*

Thank you for this comment. All scales on this screen (previously #4 and now #5) go from "not at all" to "absolutely". We made the suggested changes to make sure that we talk only about vegan diets (and not about vegan and vegetarian diets, which might create some confusion). We do not pre-register any hypotheses on these variables. We aim to discuss them as postdictions and to show some general statistics for future research. We will thus discuss the limitations if / when we introduce them in the final manuscript.

> *Inferencing*
>
> *In your power analyses, it seems like you expect a difference of 0.1 on the VDI and VPI measures, and 6 percentage points on the PMPI. Please add an explanation for why those are the expected effect sizes. There may be previous research justifying the effect sizes, or it may just be that this is what you consider practically meaningful. In all cases, I would recommend adding an explanation. It is also not clear to me if these thresholds are used for making inferences about the data, or if they are used for the purpose of the power analysis specifically. For example, if you obtain a significant difference between conditions, but the difference is less than 0.1 on a 0-1 scale, would you still infer that the intervention was effective? I am unfamiliar with using unilateral Wilcoxon tests, so I cannot comment on using it as an inferential tool. My personal preference would be to use methods that do not rely on null hypothesis significance testing (e.g., Bayesian estimation of parameter values, Bayes factors, equivalence testing), but I leave this to the editor who may be experienced with other methods.*

Thank you also for these remarks. We agree with your criticisms. We decided now to use a Cohen's d of 0.5 (usually considered as a medium effect) for the smallest effect size of interest (SESOI). We decided to rely on this arbitrary measure as there are no previous studies on which we can rely to estimate this. We also changed the estimation method and now rely on a Tobit estimation with all our outcome variables rescaled between 0 and 1.

> *Thank you very much for this interesting read.*

Thank you!