We thank the Reviewer for his most recent feedback and implemented the suggested changes. We hope that the Reviewer will now agree that the manuscript is ready for in-principal-acceptance.

1) *The justification for the relevance of the detectable effect size for each claim needs to be in the paper itself, not just in response to me;*

   We agree with the Reviewer that this might be interesting and relevant information for the reader. The justification of the relevance of the estimated detectable effect size is now added in a new section after the sample size rationale in the Supplementary Materials (pp. 25-26).

2) *The conclusions that follow need to be unambiguous. At the moment there is for som-e rows a statement about a definite conclusion given a non-significant result in the final column - but a note at the bottom disavowing the conclusion. Instead the conclusion in the final row should accurately reflect the actual conclusion legitimated by non-significance, e.g. no definite conclusion will be drawn.*

   We added a clear statement for each applicable row that we will refrain from drawing any definitive conclusions in case of non-significant results. Additionally, we removed all previously stated conclusions in case of a non-significant outcome in these rows to avoid ambiguity, since we understand that will not be able to draw conclusions on these statements given our limited sample size.

3) *Mixing inferential systems can lead to contradictions. A default BF only after a non-significant result has two problems. i) The BF may have produced evidence for H0 even if the frequentist test had been significant; ii) the prior - the model of H1 - has not been justified as relevant to the scientific problem; that meamns the BF has not been justified as relevant to the problem. The thing to do is pre-register inferences based on one system only. Report default BFs if you wish as information for the reader, but do no pre-register conclusions that follow from the Bf; instead the conclusions follow from the frequentist statistics. And then if you report BFs for some tests, report BFs for all similar tests, so there is consistency.*

   We thank the Reviewer for these more in-depths explanation of the issues he sees in the mixing of inferential systems. Accordingly, we have removed this section from the manuscript (p. 16 and Hypotheses table). We will follow the advice of the Reviewer for the stage II manuscript.

4) *post hoc power (if this means basing power on the obtained effect size) simply reflects what the p value ended up being. So don't report post hoc power. Instead decide now what the power is, and define the possible set of inferences now based on the power determined now.*

   We removed the post-hoc power calculation from the manuscript. We hope that the manuscript in its current form reflects the points that the Reviewer would like to have determined at this stage.