Response to reviewers Round 3, PCI Registered Reports #493
Please find below our reply (in blue) and the main part modified in the manuscript (in green). All changes are also reported in blue in the new version of the manuscript.

Thank you again for submitting the revised version of your Stage 1 RR to PCI RR, and for being responsive to previous comments. Most comments have been addressed satisfactorily. The R script makes the data analysis plan much more concrete, and I would recommend including it as part of your pre-registration to reduce researchers' degree of freedom in data analysis.

A redirection toward this script can now be found in the manuscript. This reads p9: "A R script demonstrating the full analysis pipeline on random data can be found at our OSF page under the "SCRIPT" folder (https://osf.io/s4trh/?view_only=4934c0215f2943cfb42e019792a30b53)".

The sample size justification is now based on resource constraints, which seems to be a more accurate reflection of the actual situation. That being said, I think important issues on sample size justification remain, and need to be further addressed.

First, since no new participants will be recruited to replace excluded ones, I assume the final sample size will be below 140. Based on your prior experiences, how many participants do you expect to exclude, and accordingly, how many do you expect to retain in the final sample? The sensitivity analysis will need to take this into account, and thus be based on the final sample size (i.e. after potential exclusions).

As described in the manuscript, end of the sampling plan section, "Because of the nature of this study, where participants are continuously recruited, some participants may still be in training after reaching the 140th complete participant, thus resulting in an eventual larger sample size". From data of our previous online studies (over 2k of recruited participants), a minimum of 15 participants that will finish the study are always in the process of training. So, when the study will reach the 140th complete participant (marking the end of recruitment), ca. 15 participants will still arrive to in the dataset, totaling to a 155. We expect to exclude 4 or 5 participants to comply with the positive controls, and 8-10 due to the exclusion of distribution outliers (total exclusion: 12 to 15). In the end, we should thus reach an estimated 140 participants after exclusion.
This rationale can now be read at the end of the sampling plan section, p4: "Because of the nature of this study, where participants are continuously recruited, some participants will still be training at the end of the recruitment phase. From previous data of our group, we expect ca. 15 participants to complete the study after the 140th, totaling to a sample size of 155. We expect to exclude 4 or 5 participants to comply with the positive controls, and 8-10 due to the exclusion of distribution outliers (total exclusion: 12 to 15). In the end, we should reach an estimated 140 participants after exclusion".

Second, and more importantly, after conducting a sensitivity analysis, you will still need to interpret the smallest effect size that can reasonably be detected, and justify why you think an investigation with such a sample size is worthwhile. After all, if a certain sample size only offers a reasonable chance to detect fairly large sample sizes, one may argue that the result will not be informative, and it is not worthwhile to embark on a project. In the manuscript, you mentioned that 140 participants (but see above) provide sufficient power to detect Cohen's d of 0.5, and r of 0.4, which are deemed 'relevant' and 'non-negligible'. Most people would probably agree that these two effect sizes are relevant in the current context. However, personally I would consider a smaller effect size for H1, e.g. Cohen's d of 0.3, also to be relevant, but 140 participants only provide around 55% power to detect this effect. The question that needs to be better addressed is thus why you think this investigation is still worthwhile, given that it will likely miss smaller but potentially also relevant effects. (In other words, why is it okay to miss smaller effects in the current investigation?)

To be clear, I am not asking for a 'general-purpose justification'. The interpretation of effect sizes is supposed to be highly context-dependent. The reasons that you mentioned in the response letter seem relevant. In the manuscript, you mentioned the sample size to be sufficient, without really saying ***why***. This 'black box' needs to be open, so that readers can better judge the soundness and validity of the justifications. This is a rather important issue - sample size justification based on resource constraints need to be subject to the same level of scrutiny as other methods (e.g., a priori power analysis), to ensure that the results will be informative. I sincerely hope that this issue can be sufficiently addressed in this round of revision.
Kind regards,
Zhang Chen

Since the recommender considers relevant the justifications for the SESOI we included in the Round 2 response letter, we have now reported the main ones in the manuscript. It reads now, p4 and 5:
"For H1, […] an additional 5 days of diet (extracted from a Cohen's d of 0.5 with an estimated standard-deviation of 10 days) would be associated with physiological and cognitive modifications that might be detectable and considered relevant by the participants and the health care providers (i.e., reduction in appetite, higher energy level stability, induction of consumption habits, and realization by the participant that restriction can be maintained).

For H2 and H3, […] Because correlations capture both causal relationships and indirect connections, the observed correlations in our study will inherently exceed their causal effects. If we were to identify correlations below 0.4 for both H2 and H3 (equivalent to 16% of explained variance), it would signify that less than 16% of the variance is attributable to causation. This criterion is the lowest that we consider ensuring that our findings effectively justify to conduct further research on these relationships' (causal) significance.

While impactful effects of restriction would need longer reduction of sugar intake to take place (reduction in weight, dental health improvement, reduced risk of non-alcoholic fatty liver disease, etc.), we consider that reaching 5 additional days of restriction would represent a proof of principle that MIT interventions can facilitate restrictive diets. Likewise, we consider the indication for the correlative association we target between the devaluation, amount of training and days of successful dieting to be minimally sufficient to justify trials testing a causal association between these factors. We acknowledge that smaller effect sizes could also be relevant, but we set these large smallest effect sizes of interests to reinforce the argument to conduct on this basis heavier interventional research efforts."