**Review for Stage 1 RR:**

**"Sugary drinks devaluation with executive control training helps to resist to their consumption"**

***Sampling plan***

If 50 participants are needed in the experimental group for H2 and H3 why is the target sample size based on H1 which requires a smaller number of participants in each group (N= 36)? Also, I would be more inclined to use a *d* of 0.4 for the power analysis and not the chosen mean/sd differences that result in a Cohen's *d* of 0.7, as in the rest of the manuscript you mention a *d* of 0.4 as your benchmark (smallest effect size of interest). In any case please make this clearer as reading this the first time I thought that the 7-day difference resulted in a *d* of 0.4. If you apply this change, your total sample size would be 216. A greater sample size may also allow for more power for H2 and H3 analyses (under 50 participants in the experimental group - which will further be reduced if you apply data exclusions as per my comment for the *Pre-post explicit liking reduction* section).

***Recruitment and screening***

As an inclusion criterion, willingness to follow a restrictive diet is important but it is also worth recording participants' baseline consumption behaviour. Are you including everyone from people who rarely drink sugary drinks to people who drink more than a few sodas a day? In this section it is important to also add any methodological details or at least point to where the reader can access them (what are the exact questions for your screening- e.g. how do you define 'healthy' individuals in this context).

***Training tasks***

I understand that the tasks have been used in previous studies but for this Registered Report I think you should not omit the methodological details and specific parameters of the tasks being administered (contingencies, time limits, number of trials, feedback element etc.). They are central to the study and should be presented as part of the Stage 1 proposal for further evaluation - even if the app cannot be changed at this time. The video for the app demonstration was really helpful - you could add a figure with screenshots from the game in the main text for convenience as well.

***Questionnaires***

Please add a reference to supplementary material or an online repository where the *full* questionnaires can be found. In the text you can add more details about the 10 items being included in the health questionnaire - what are you measuring with regards to participants' health?

***Analysis plan***

In this section you need to specify all the analyses that will be run and treated as 'confirmatory' with details so you can move the paragraphs from the *Statistical contrasts* section here. For Bayesian analyses, what priors will you be using for the t-test and correlations? Also, while BFs can be very informative in the case of inconclusive results it would be preferable in my opinion to report them for all results irrespective of significance. Also, although it may seem obvious please state on which statistics you will base your conclusions on (e.g. frequentist but BFs reported in a supplementary manner?).

***Data exclusions***

Could you add details here about potential missing data and related exclusions with regards to your questionnaires? For example, what if participants don't complete the weekly questionnaires, or if information is missing (e.g. exact dates of first consumption etc.)?

Given that reaction times from the analogue scales are recorded I presume that you can also have access to training performance data. While technically a day of training can be counted as successful if completed, I believe you should mention whether potential exclusions can apply to adherence. For example, does training proceed if you miss the reaction time window (if there is one, not currently known based on the details presented) or if participants simply skip trials and not interact with the game?

Since the proportion of successful inhibitions in such training tasks may be a moderator of training effects (see meta-analysis by Jones et al. 2016), it is worth considering a performance benchmark for data exclusions - e.g. if one day of training is completed but participants fail to stop for more than half of the trials. As this may be a conservative criterion for data exclusions given the sample size you propose, it would be interesting to add

training performance as a secondary outcome or consider certain exploratory analyses at Stage 2 to look at learning effects and inhibition success.

***Statistical contrasts***

Please state why have you chosen this criterion for your effect interpretation - i.e. why is a min *d* of 0.4 required to consider the result 'relevant' - I have found this confusing as mentioned in the *Sampling plan*. You mention that the result will only be relevant if the difference is at least 7 days or more of successful dieting, but in your power analysis this corresponds to a Cohen's *d* of 0.7 and yet in the text a Cohen's $d \geq 0.4$ is treated as relevant.

***Baseline reported consumption***

If I understood this correctly, will you inspect the data, run the analyses and if for H1 you get a Cohen's *d* greater than 0.4 you will exclude participants and report the results with the reduced sample size? For other data exclusions that do not require statistical analyses I assume that recruitment will continue until the sample size target is met, but for this exclusion criterion please add more details regarding the sampling plan - that is, how your sample size may be affected given that the target is based on an *a priori* power analysis.

***Pre-post explicit liking reduction***

Please add a justification for this in a narrative format - e.g. for H2 you only want to run the correlations if a devaluation effect is observed (defined by your chosen threshold) - this should be clear in your hypotheses as well. It may be good to present results without all the effect-related exclusions in the supplementary material as well for comparison.