PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Acute and repeated impact of sweeteners and sweetness enhancers in solid and semi-solid foods on appetite - protocol for a multi-centre, cross-over, RCT in people with overweight/obesity: The SWEET Project
AUTHORS	Gibbons, Catherine; O'Hara, Beverley; O'Connor, Dominic; Hardman, Charlotte; Wilton, Moon; Harrold, Joanne; Almiron-Roig, Eva; Navas-Carretero, Santiago; Hodgkins, Charo; Nazare, Julie Anne; Alligier, Maud; Martínez, Jose; Scott, Corey; Kjølbæk, Louise; Normand, Mie; Rannou, Cécile; Blaak, Ellen; Feskens, Edith; Moshoyiannis, Hariklia; Raben, Anne; Halford, Jason; Beaulieu, Kristine; Finlayson, G

VERSION 1 – REVIEW

REVIEWER	Graybeal, Austin J University of Southern Mississippi
REVIEW RETURNED	31-May-2022

GENERAL COMMENTS	This study seeks to investigate the influence of sweeteners and sweetness enhancers on appetite and food intake compared to a control in individuals with overweight and obesity across multiple sites in Europe. While this manuscript seeks to simply preregister their methods, the rationale for conducting the study requires major revisions and the methodology, minor revisions. These are described line by line below and I look forward to reviewing these revisions. Line 79: Use of the term “political environment” is too broad a term to associate with drivers of obesity. I am also not certain that the reference used here supports that claim. Wouldn’t “socioeconomic status”, “economic environment”, or “public (nutritional) policy” be better here? Line 80: I understand what you are trying to say, but is it worth the words to discuss physical activity only to never come back to it for your rationale? Would a better way to say this be “….diet too rich in energy intake relative to energy expenditure”? Line 81-92: This sections seems to jump out of nowhere and would benefit from an improved transition (i.e., why I believe that the physical activity language should be removed). Given our knowledge on the appetite response to carbohydrates (sugars) relative to other macronutrients, and that appetite is the primary endpoint in your study (Line 52), would it not be better to use the appetite response to free-sugars/carbohydrates in general as the rationale for your study and spend more time describing how S&SE could be a solution or worsen the existing problem (Line 88-92). You could also discuss why low-sugar diets have been

	unsuccessful (cravings, palatability, acceptance, familiarity, etc.) which would transition well into line 88. Line 87: This sentence “This is a recommendation…” is contradicted by lines 118-120. Is sugar intake being addressed in Europe (through public policy or commercial products) or is it not? One of these sections should be removed. Removing all of the sugar intake recommendations from the WHO in the first paragraph would improve the manuscript, in my opinion, and thus there would be no need for line 87. This would allow for an initial paragraph more focused on the study rationale (i.e., appetite responses to carbohydrates and sugar). Line 94-97: I think it would improve the manuscript to collapse these sentences into one. Lines 96-97 are a bit confusing given that you do not describe their “modes of action” which can be mean many different things (neurological, pharmakinetics, physiological, neuropsychological). You could say, “….types of studies, comparisons, and S&SEs being used”. Line 111-122: The readability is not where it needs to be although the information is sufficient. I think you could condense this entire paragraph into the above paragraph to increase the clarity. For instance, you could move lines 111-114 after Line 95 where you begin the literature review. You could also move line 115-122 after line 106 and delete line 107-109. It is my opinion that it would improve the manuscript and overall “get to the point” in an organized manner. Line 128: Since you stated earlier that S&SEs have different modes of actions (presuming this would also be different than a control in a positive or negative direction) would you not have a directional hypothesis here? It might be better to hypothesize that there will be differences from the control and direction (differences from the control) will vary by S+SE matrix. Same for the following hypothesis. Line 143: I know that logistically, this may be difficult, but would the intervention not be improved by adding a fully liquid condition (i.e., a beverage). There is obvious application here, but more so, you would be able to show the differences across the spectrum of food viscosity and thus create a more comprehensive study. It would reduce the need to state that in the limitations “we were unable to see/compare the responses in solid and semi-solid foods to beverages”. I understand the rationale for the food matrices you selected, but if you would be willing to supplement one of those conditions with a beverage it would provide an extra layer to the study that it is currently missing. Maybe the study subjected to the COVID restrictions (i.e., the cake matrix) should be replaced with a beverage condition since you cannot collect the important blood data for that condition anyways? Or replace the cake matrix with the beverage, but reimplement the cake matrix at a setting allowed to collect blood markers since the cake condition is pivotal to the design (solid vs semi solid foods). Line 159: Why not an alpha of 0.05 as used for the other outcomes? Line 162: Why was peak pancreatic polypeptide at 15 min used for the sample size in the study when the investigation overall is seeking to primarily examine iAUC? I see that the referenced study used beverages across different levels of viscosity (similar to your design), but using the 15 min timepoint is odd. If you chose this due to similar study designs and because you expect responses to be observed in the blood at 15 min postprandial, please state it and why it is more useful to use a single timepoint compared to the summed response. I think this is important to

	think through because you will ideally see that hormonal appetite responses to the S&SE will be blunted compared to the control (PMID: 21255472), but that this may not result in similar responses to subjective appetite responses. Thus, you would be powering to a result that without much clinical significance from an eating behavior perspective. Line 166: What about for GLP-1 and insulin? Describe how we would assume that the sample was sufficient across all peptide outcomes used in the study. Without describing the other hormones, how would the reader know that you powered to the lowest effect size of interest. I.e., did you choose PP because GLP-1 and insulin had smaller effect thus requiring a larger sample? Or is their physiological rationale/lack of existing data that supports the exclusion of other markers? Line 249: Should high-intensity physical activity not require a longer abstention period? It would be safer to instruct participants to abstain from exercise for 24 h prior to the CIDs to remove the influence of exercise on the appetite related outcomes. Given your sample and exclusion criteria, I do not (personally) foresee that being an issue. Line 266-267: Why not supply each group with a tablet to collect this? Or only perform the assessment on a PC across all groups? It is odd to use different devices, although valid, when both devices are so widely used across universities. Line 281: This is simply a recommendation to the researchers based on some of my previous perceptions conducting this type of research: Would it be advantageous to standardized what the participants are doing between measurements both between and within subject? I agree that no food-related stimuli is a must, but what if the participants decide to watch a horror-movie, or read aggressive twitter threads, etc.? Would these not influence something like desire to eat that varies day-to-day? Or, for instance, if you allow student to work on assignments or non-students to perform work-related tasks on a computer, would assignment/task deadlines, for example, not enact varying stress responses that could influence appetite? In my experience, setting up a computer or tablet to watch something like Netflix (with restrictions being shows/movies that are not food-related/horror) can control the between participants aspect. You can control the within participants aspect by keeping them within a genre of their initial choosing across conditions. Thus, they are receiving the same stimuli each time. If you choose not to do this, implementing a mood questionnaire at each CID would be advantageous even if it is only used as a covariate. Line 406: You have several nominal variables used as covariates but state that you will only adjust when these covariates “correlate” with outcome variables. Please describe the non-parametric tests you will use to verify correlations between continuous and discrete variables, if that is how you plan to rationalize your adjustments.

REVIEWER	Nadolsky, Karl Spectrum Health, Endocrinology
REVIEW RETURNED	10-Jun-2022

GENERAL COMMENTS	Protocol for acute and repeated dose impact of sweeteners and sweetness enhancers on appetite-related behaviour, physiology, and health: a multi-centre, double-blind, crossover, randomised, controlled trial in people with overweight/obesity. The SWEET project.

	Suggest specifying, “non-nutritive” sweeteners (NNS) Any inclusion/exclusion criteria based upon cardiometabolic complications of obesity beyond BMI? Such as insulin resistance, metabolic syndrome, prediabetes, type 2 diabetes, etc? Their inclusion would help note glycemic effects but may not be powered for such CGM data would be helpful

REVIEWER	Higgins , Kelly USDA Agricultural Research Service
REVIEW RETURNED	27-Jun-2022

GENERAL COMMENTS	The current protocol addresses an interesting research question regarding the effect of Sweeteners and sweetness enhancers (S&SEs) on appetite and related behavioral, metabolic and health outcomes. The authors provide adequate background to justify their desire to study this research question. While I have concerns regarding the duration and sample size of the study protocol, I believe that these may not be modified given the manuscript suggests the research is currently underway. Additional information regarding the rationale for certain decisions can be provided which enhance the interpretation of the results of this research and will help other authors design future research questions. Specific revisions are included in the attached word document.

VERSION 1 – AUTHOR RESPONSE

Reviewer Reports:

Reviewer: 1
Dr. Austin J Graybeal, University of Southern Mississippi

Comments to the Author:
This study seeks to investigate the influence of sweeteners and sweetness enhancers on appetite and food intake compared to a control in individuals with overweight and obesity across multiple sites in Europe. While this manuscript seeks to simply preregister their methods, the rationale for conducting the study requires major revisions and the methodology, minor revisions. These are described line by line below and I look forward to reviewing these revisions.

We thank the reviewer for their time and attention given to this protocol paper and for the improvement these changes will bring. For context, the wider SWEET project (https://sweetproject.eu/) is a collaborative project involving 27 different partners across the EU. The project has been funded contingent on a contractual grant agreement based on the Description of Work as proposed by the consortium to the European Commission and data collection has started at most intervention centres. Therefore, we have some constraints preventing us from changing major aspects of the design and methodology. This means that we were not able to action all of the proposed changes by the reviewer, but where applicable (and noted below) we have noted these issues in the manuscript (please note line numbers noted below are from the clean version of the manuscript). We especially appreciate the valuable feedback on the study rationale and logic behind

the sample size determination, which have considerably improved the manuscript.

Line 79: Use of the term “political environment” is too broad a term to associate with drivers of obesity. I am also not certain that the reference used here supports that claim. Wouldn’t “socioeconomic status”, “economic environment”, or “public (nutritional) policy” be better here?

Thank you for this recommendation, we have changed the wording to ‘public (nutritional) policy’ (line 79).

Line 80: I understand what you are trying to say, but is it worth the words to discuss physical activity only to never come back to it for your rationale? Would a better way to say this be “….diet too rich in energy intake relative to energy expenditure”?

Thank you for the suggestion, we have made the recommended change (line 80).

Line 81-92: This sections seems to jump out of nowhere and would benefit from an improved transition (i.e., why I believe that the physical activity language should be removed). Given our knowledge on the appetite response to carbohydrates (sugars) relative to other macronutrients, and that appetite is the primary endpoint in your study (Line 52), would it not be better to use the appetite response to free-sugars/carbohydrates in general as the rationale for your study and spend more time describing how S&SE could be a solution or worsen the existing problem (Line 88-92). You could also discuss why low-sugar diets have been unsuccessful (cravings, palatability, acceptance, familiarity, etc.) which would transition well into line 88.

Thank you for this suggestion. We have adjusted the rationale to incorporate this relevant argument which fits better with our study outcomes.

Line 87: This sentence “This is a recommendation…” is contradicted by lines 118-120. Is sugar intake being addressed in Europe (through public policy or commercial products) or is it not? One of these sections should be removed. Removing all of the sugar intake recommendations from the WHO in the first paragraph would improve the manuscript, in my opinion, and thus there would be no need for line 87. This would allow for an initial paragraph more focused on the study rationale (i.e., appetite responses to carbohydrates and sugar).

Thank you for pointing this out, we have removed the WHO recommendations and agree that this aids the flow and clarity of the manuscript.

Line 94-97: I think it would improve the manuscript to collapse these sentences into one. Lines 96-97 are a bit confusing given that you do not describe their “modes of action” which can be mean many different things (neurological, pharmakinetics, physiological, neuropsychological). You could say, “….types of studies, comparisons, and S&SEs being used”.

Thank you for this suggestion, we have reduced to one sentence.

Line 111-122: The readability is not where it needs to be although the information is sufficient. I think you could condense this entire paragraph into the above paragraph to increase the clarity. For instance, you could move lines 111-114 after Line 95 where you begin the literature review. You could also move line 115-122 after line 106 and delete line 107-109. It is my opinion that it would improve the manuscript and overall “get to the point” in an organized manner.

We have made these suggested changes in the manuscript and agree this improves the introduction. We retained lines 107-109 as we feel this is a key component of the design of the current study.

Line 128: Since you stated earlier that S&SEs have different modes of actions (presuming this would also be different than a control in a positive or negative direction) would you not have a directional hypothesis here? It might be better to hypothesize that there will be differences from the control and direction (differences from the control) will vary by S+SE matrix. Same for the following hypothesis.

While S&SE are heterogeneous in terms of chemical structure, organoleptic properties and physiological effects, the evidence on the direction of effects on appetite-related outcomes is insufficient to make specific predictions for each S&SE blend. The findings of our study may lead to directional hypotheses in future research.

Line 143: I know that logistically, this may be difficult, but would the intervention not be improved by adding a fully liquid condition (i.e., a beverage). There is obvious application here, but more so, you would be able to show the differences across the spectrum of food viscosity and thus create a more comprehensive study. It would reduce the need to state that in the limitations "we were unable to see/compare the responses in solid and semi-solid foods to beverages". I understand the rationale for the food matrices you selected, but if you would be willing to supplement one of those conditions with a beverage it would provide an extra layer to the study that it is currently missing. Maybe the study subjected to the COVID restrictions (i.e., the cake matrix) should be replaced with a beverage condition since you cannot collect the important blood data for that condition anyways? Or replace the cake matrix with the beverage, but reimplement the cake matrix at a setting allowed to collect blood markers since the cake condition is pivotal to the design (solid vs semi solid foods).

We thank the reviewer for making this interesting point. In fact, there is another body of preliminary work within the wider SWEET project that is examining S&SE in beverages. This study used some of the same S&SEs as selected in the semi-solid and solid food matrices. While time and budget prevents us from adding another condition to the present study, it will be interesting to compare the findings of this study to the outcomes of the beverage study which is currently being written up for publication. We have added 'semi-solid and solid food matrices' to the title of the paper to aid clarity.

Line 159: Why not an alpha of 0.05 as used for the other outcomes?

We apologise for the inconsistencies in approach for the sample size calculations. We have addressed this issue under your next point 2 points and made relevant changes in the manuscript.

Line 162: Why was peak pancreatic polypeptide at 15 min used for the sample size in the study when the investigation overall is seeking to primarily examine iAUC? I see that the referenced study used beverages across different levels of viscosity (similar to your design), but using the 15 min timepoint is odd. If you chose this due to similar study designs and because you expect responses to be observed in the blood at 15 min postprandial, please state it and why it is more useful to use a single timepoint compared to the summed response. I think this is important to think through because you will ideally see that hormonal appetite responses to the S&SE will be blunted compared to the control (PMID: 21255472), but that this may not result in similar responses to subjective appetite responses. Thus, you would be powering to a result that without much clinical significance from an eating behavior perspective.

Gut hormones are a secondary outcome of the studies so they were not included in the main power calculation. Nevertheless, to confirm that there was a chance to see differences between S&SE products and the control product, the minimum sample size for gut hormones was estimated based on observed values for glucose, insulin and Pancreatic Polypeptide according to Yeomans et al. Int J Obes 2016. This study was chosen for its similar cross-over design with serial blood samples (N=8)

taken over 90 minutes using beverages that differed across levels of energy and viscosity (i.e. energy containing foods). There were also some differences, in that only men were included. Therefore, in consultation with statisticians at 2 intervention centres, we estimated that a minimum of 30 completers per matrix would be adequate for these secondary outcomes with 80% power, alpha 0.05 and a strong (0.9) effect size.

Our statement that minimum sample size was based on the 15' peak for PP was a misunderstanding by one of the authors which we thank the reviewer for correcting. Due to the interesting finding by Yeomans et al. that PP levels returned to baseline within 15 minutes in low viscosity conditions, we included a measure of PP at the 15 minute timepoint to increase our chance of capturing the peak in our iAUC calculation. The 15 minute peak was not used as the basis of the power calculation as stated and we apologise for this error.

We have included the changes to the manuscript under the next point and we appreciate the opportunity to improve the clarity of this section.

Line 166: What about for GLP-1 and insulin? Describe how we would assume that the sample was sufficient across all peptide outcomes used in the study. Without describing the other hormones, how would the reader know that you powered to the lowest effect size of interest. I.e., did you choose PP because GLP-1 and insulin had smaller effect thus requiring a larger sample? Or is their physiological rationale/lack of existing data that supports the exclusion of other markers?

To our knowledge, there are no papers that follow a similar design and measure the same combination of peptides as our present study. Yeomans et al. study was chosen due to the similar design to our study using energy containing foods that differed in energy and sensory properties, and that it measured a number of peptides within the study. They measured glucose, insulin, PP and CCK and found sensory effects or interactions with PP and insulin using 23 participants. Another study (Bowen et al. 2006) comparing carbohydrate and protein preloads suggested a similar sample size would be sufficient to detect a strong effect size for total ghrelin. Another study (Gibbons et al, 2013) comparing fat and carbohydrate preloads measured glucose, insulin, ghrelin, GLP-1 and PYY and found differences in all except ghrelin with a sample size of 16.

We have amended the manuscript lines 147-161.

Line 249: Should high-intensity physical activity not require a longer abstention period? It would be safer to instruct participants to abstain from exercise for 24 h prior to the CIDs to remove the influence of exercise on the appetite related outcomes. Given your sample and exclusion criteria, I do not (personally) foresee that being an issue.

We agree this relatively short abstention period should not be an issue for this study. We also have an exclusion criteria that excludes participants who do more the 10 hours vigorous activity per week.

Line 266-267: Why not supply each group with a tablet to collect this? Or only perform the assessment on a PC across all groups? It is odd to use different devices, although valid, when both devices are so widely used across universities.

This was due to the fact that different sites have access to different devices. The UK sites, for example did not have access to enough tablets, whereas the Spanish centre did not have the set-up to be able to use PC's. Each site chose which device best suited their laboratory set up. Intervention site will also be controlled for in the statistical analyses which should reduce the potential impact of this. Furthermore, all computer systems were designed and tested on both PC's and tablets therefore we do not see any issue with this.

Line 281: This is simply a recommendation to the researchers based on some of my previous perceptions conducting this type of research: Would it be advantageous to standardized what the participants are doing between measurements both between and within subject? I agree that no food-related stimuli is a must, but what if the participants decide to watch a horror-movie, or read aggressive twitter threads, etc.? Would these not influence something like desire to eat that varies day-to-day? Or, for instance, if you allow student to work on assignments or non-students to perform work-related tasks on a computer, would assignment/task deadlines, for example, not enact varying stress responses that could influence appetite? In my experience, setting up a computer or tablet to watch something like Netflix (with restrictions being shows/movies that are not food-related/horror) can control the between participants aspect. You can control the within participants aspect by keeping them within a genre of their initial choosing across conditions. Thus, they are receiving the same stimuli each time. If you choose not to do this, implementing a mood questionnaire at each CID would be advantageous even if it is only used as a covariate.

We thank the reviewer for this interesting point. We feel that individual preference on how to spend their time is highly variable therefore allowing them to spend the time as they wish (but with the given constraints) best suited the study design given that they have to give up several hours of their time for this study. Across the consortium we have good experience in appetite and eating behaviour studies and feel this was the best option in order to aid recruitment and retention of participants in the study. We do have a measure of mood from the COEQ questionnaire on each CID therefore it will be possible to assess mood across the CIDs.

Line 406: You have several nominal variables used as covariates but state that you will only adjust when these covariates “correlate” with outcome variables. Please describe the non-parametric tests you will use to verify correlations between continuous and discrete variables, if that is how you plan to rationalize your adjustments.

We apologise for the confusion caused by our statement regarding covariates. We will not be performing correlations with covariates before performing the linear mixed models but rather will add them to our models to see if they influence outcomes. Covariates with significant effects on outcomes will remain in the models and will be reported accordingly. We have amended the text to reflect this (lines 412-414).

Reviewer: 2
Dr. Karl Nadolsky, Spectrum Health

Comments to the Author:
overall great study concept and design to continue investigating more nuanced effects of nonnutritive sweeteners.

We thank the reviewer for their positive evaluation and comments on the manuscript. We have addressed the comments in order below and pointed out where changes have been made in the manuscript.

Few thoughts:
- Suggest specifying, “non-nutritive” sweeteners (NNS)

Thank you for this suggestion. Our use of the term Sweeteners and Sweetness Enhancers (S&SE) was due to this specific term being used by the European Commission in the call text for the study grant. This is also the term we have used in another SWEET work package protocol (“bmjopen-2022-

061075 - "Protocol for a multicentre, parallel, randomised, controlled, trial on the effect of sweeteners and sweetness enhancers on health, obesity and safety in overweight adults and children. The SWEET project."). We have now clarified in the manuscript that the term S&SE refers to "non-nutritive sweeteners and sweetness enhancers" (lines 41 and 88-89).

- Any inclusion/exclusion criteria based upon cardiometabolic complications of obesity beyond BMI? Such as insulin resistance, metabolic syndrome, prediabetes, type 2 diabetes, etc? Their inclusion would help note glycemic effects but may not be powered for such

Yes, these are noted in detail in supplementary material 1, and yes, those with diabetes were excluded.

- CGM data would be helpful

We agree with the reviewer, thank you for this point. Although this was not part of the original proposal which was funded, we have since added a supplementary sub-study at the University of Leeds UK site only, and only in a smaller number of participants (n=10).

Reviewer 3

The current protocol addresses an interesting research question regarding the effect of Sweeteners and sweetness enhancers (S&SEs) on appetite and related behavioral, metabolic and health outcomes. The authors provide adequate background to justify their desire to study this research question. While I have concerns regarding the duration and sample size of the study protocol, I believe that these may not be modified given the manuscript suggests the research is currently underway. Additional information regarding the rationale for certain decisions can be provided which enhance the interpretation of the results of this research and will help other authors design future research questions. Specific revisions are included in the attached word document.

We thank the reviewer for their insight into this work and providing a number of important issues to consider for the consortium. We have dealt with your points in turn, and detailed how we have made changes to the manuscript (if applicable). Please note line numbers refer to the clean version of the manuscript.

- Are the 5 RCTs the same protocol tested at 5 different sites or 5 RCTs, each completed at 5 different sites. This needs to be clarified. Also, it is unclear what measures will be collected at what site and what the final sample size is at each site. In addition, please describe why all samples were not collected at each site and why the sample size will vary for each site?

We thank the reviewer for pointing out that this wasn't clear. In lines 168-176 we have clarified that each of the 5 sites will complete one/two RCTs following the exact same protocol but with different food matrices. The reason why all sites are not collecting all samples is due to changes that were required after the COVID-19 pandemic which meant time and research funds were limited. This has affected the cake study where there will be no blood samples and a reduced sample – we have made clear that the results of this trial will be descriptive only for this reason.

- The total sample size from the 5 sites is 213 subjects. Each site with introduce more variability that will need to be controlled for. Is this sufficient sample size total and per site?

Yes, in the full studies with 53 participants we have accounted for additional variability from sex, age and intervention site.

- The sample size determination section is lacking substantial critical information. In general, more information is needed to justify each factor included in the sample size determination. For example, how was 8 mm selected on a VAS scale determined for the calculation of the primary endpoint? Was the variability across study sites accounted for in the power calculation?

To detect a minimum difference of 8 mm in appetite ratings on a 100 mm VAS with 80% power, alpha 0.05, and based on a published within-subject SD of 14.4 mm in VAS measures (Ref 1), an overall sample of 53 subjects (both genders, across each matrix) would be needed (Ref 2, p.30). We

expect this sample size will be sufficient to detect iAUC differences in the appetite response relative to control of around 8–10%, considered to be of practical relevance (Ref 3).

Ref 1. Almiron-Roig E, Green H, Virgili R, Aeschlimann JM, Moser M, Erkner A. Validation of a new hand-held electronic appetite rating system against the pen and paper method. Appetite. 2009;53(3):465-468.
Ref 2. Jones B, Kenward MG. Design and Analysis
Ref 3. Blundell, J., De Graaf, C., Hulshof, T., Jebb, S., Livingstone, B., Lluch, A., Mela, D., Salah, S., Schuring, E., Van Der Knaap, H. and Westerterp, M., 2010. Appetite control: methodological aspects of the evaluation of foods. Obesity reviews, 11(3):251-270.
We have added further clarity to the manuscript reflecting the above (lines 149-161).

- In the strengths/limitations, changes to the protocol due to the COVID-19 pandemic are noted. However, they are not discussed at any other point in the protocol. Why where these changes to the protocol necessary and how will they affect the interpretation of the results?

Yes, this is a good point. It is briefly mentioned under the sample size calculations but we have added further information in the 'Trial Status' section of the manuscript (lines 439-453). We believe the interpretation of the results will be unaffected if the sample size meets the power calculation. For the smaller studies the findings will be presented descriptively and interpreted with extra caution.

- Most sweeteners are consumed as beverages, and as the authors stated, most studies are conducted using beverages. Why was a beverage matrix not selected? This would serve as a beneficial treatment and potentially a good control to compare to results from other previously conducted studies.

We thank the reviewer for this suggestion. As mentioned above in the response to reviewer 1, another preliminary study within the wider SWEET project was performed using beverages, therefore while a beverage condition is not feasible in the current study design, it will be interesting to compare the findings of this study to the outcomes of the beverage study which is currently being written up for publication. We have also specified in the title that this study uses semi-solid and solid food matrices.

- Lines 83-84 "and its potential to add to the overall energy density of diets". Sugars are carbohydrates and have generally low energy density compared to other macronutrients (e.g. fat). In addition, most sugars are consumed in the form of beverages, which have an innately low energy density. This statement is flawed.

Thank you for pointing this out. The statement has been amended.

- Lines 94-109 The preceding sentence states that the impact of S&SE on appetite and health related outcomes is unclear. I agree with that statement, but the following paragraph seems to suggest a relatively clear benefit of the use of sweeteners, at least with regard to energy intake. I think this paragraph could benefit from providing more information about why this effect is "unclear."

We have reworked this paragraph to make the overview of the literature clearer based on some comments by reviewer 1. We believe this improves the clarity of the manuscript in its revised form.

- Line 202 Whether energy is substituted, added, or replaced is a very important concept when studying sweeteners. How the sweeteners are incorporated into the diet should be discussed further and should provide a rationale for this decision.

Thank you for pointing out this omission. The text has been amended accordingly with the following information regarding the substitution strategy (lines 200-205).

- Lines 226-228 "an eligibility taste test of the control intervention product." Please describe this test further.

Yes, we agree this is important information and have added the details to the manuscript (lines 226-228).

- Figure 1. All abbreviations should be defined. Provide information regarding the contents of the specific products in the footnotes.

The abbreviations have been defined in figure 1. With regards to the contents of the products we have included this in supplementary material 2 which is referred to in the relevant section of the manuscript.

- Lines 303-304 "Intervention and control products will be matched for sweetness intensity, flavour and physical appearance." Describe how this will be accomplished

The producers of the food products have completed this testing during the formulation of the products using a sensory panel approach. In addition, similarity of taste and palatability between the product conditions will be verifiable from the data collected in the current study.

- Lines 306-307 Provide rationale for how/why these sweeteners and blends were selected. In generally information regarding the actual sweeteners and enhancers used is lacking, especially given how the differences between sweeteners is highlighted in the introduction.

Yes, we agree that more information is needed, thank you for pointing this out. The sweeteners and blends were selected based on the available literature (neotame and Stevia Reb M, as examples of 'artificial' and 'natural' sweeteners, were selected for biscuits and cakes) and also based on the results of the preliminary beverage study (Stevia RebM and Mogroside V; Sucralose and Acesulfame K were selected for the yoghurt, chocolate and cereal). The results of the beverage trial are being written-up for publication and we will be able to cite these results as a basis for the selection in due course. Information regarding which S&SE / blends included in each product has been added to the manuscript (lines 307-311).

- Lines 404-405 "Analyses will include repeated measures ANOVA or mixed-effects regression models" Why is this decision not made a priori?

ANOVA models were initially considered in the event that no missing data would be encountered; however, in the case of blood samples, for example, to account for missing samples, linear mixed models are more appropriate. We have since decided to use linear mixed models for all analyses for consistency. This is now reflected in the manuscript (lines 408-415).

- Lines 406-408 "Covariates... will be adjusted for where necessary (e.g., if they correlate with other variables)." This is not sufficient information. What level of correlation? Analyses should be reported as both unadjusted and adjusted for completeness of data reporting.

We apologise for the confusion caused by our statement regarding covariates. We will not be performing correlations with covariates before performing the linear mixed models but rather will add them to our models to see if they influence outcomes. Covariates with significant effects on outcomes will remain in the models and will be reported accordingly. We have amended the text to reflect this (lines 412-414).

VERSION 2 – REVIEW

REVIEWER	Graybeal, Austin J University of Southern Mississippi
REVIEW RETURNED	29-Aug-2022

GENERAL COMMENTS	The authors have sufficiently addressed all of my previous comments and the comments from other reviewers and thus, the manuscript is acceptable for publication. Best of luck to the research team on their efforts and I look forward to the publication of these works.

VERSION 2 – AUTHOR RESPONSE

We than the editor for the opportunity to make these amendments and hope the manuscript is now ready for acceptance.

- The abstract >> ethics and dissemination section needs to make it clear that this study has received ethics approval (we do not publish protocols that have not received ethics approval).

We thank the reviewer for pointing this out. Four out of five of the study centres have received ethical approval and the fifth centre are currently awaiting a final outcome, which is due soon. We have amended the abstract to make this clear.

- The 'ethics and monitoring' section in the main manuscript also indicates that you are yet to receive ethics approval for this research. Can you please confirm that approval has been obtained from the ethics committees of each participating site? Please provide the specific names of the ethics committees that have approved the study along with the approval reference numbers. If approval has yet to be obtained from some sites than can you please clarify when you expect to receive approval? We ideally would only publish the protocol after ethics approval has been obtained from all centres.

The text is this section has been amended to the following:

The following details the specific ethical committees and the reference numbers. University of Leeds School of Psychology (PSC-127, approved 19th November 2020), Comité de Protection des Personnes Nord-Ouest III (2021-42, approved 28th March 2022), Comité de Ética de la Investigación de la Universidad de Navarra (2021.205, approved 7th March 2022), The Ethical Committee, Region H. Denmark (application number H-21078447, submitted and have made comments back to review committee, final decision is imminent) and University of Liverpool Central University Research Ethics Committee D (10659, approved 14th April 2022). All study procedures will be conducted in accordance with the Helsinki Declaration and the study protocol has been registered in a public database (clinicaltrials.gov NTC04633681; Supplemental Material 6).

- Page 6: Please can you remove 'bmjopen-2022-061075' from the manuscript and instead cite the companion protocol? The reference can be amended during the proofing stage if the two papers are accepted for publication.

Yes, this has been removed as per the editor's suggestion.