PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Protocol for the Gut Bugs Trial: a randomised double-blind placebo-controlled trial of gut microbiome transfer for the treatment of obesity in adolescents
AUTHORS	Leong, Karen; Jayasinghe, Thilini; Derraik, José; Albert, Benjamin; Chiavaroli, Valentina; Svirskis, Darren; Beck, Kathryn; Conlon, Cathryn; Jiang, Yannan; Schierding, William; Vatanen, Tommi; Holland, David; O'Sullivan, Justin; Cutfield, Wayne

VERSION 1 - REVIEW

REVIEWER	Jed Friedman University of Colorado School of Medicine, USA
REVIEW RETURNED	25-Sep-2018

GENERAL COMMENTS	-Can you comment further on the primary outcome on which the study is powered? On p. 12 it indicates a BMI difference of 0.19 at 6 wks -this seems small and would be subject to less clinically meaningful as say DEXA, OGTT, or a change in lipid levels. -Secondary outcomes beyond food questionaires including satiety measures might be important. As the MB may only provide a small short-term effect on body weight unless diet is modified, this would seem to be more important to assess outcomes. SCFA are currently in vogue for insulin sensitivity and for metabolic effects and should be considered. The central question beyond body weight are what determines whether this study can reveal important functional elements of the microbiome such as (i) is the observed variability biologically meaningful?, and (ii) is a measured microbial functional state identifiable? For the metagenome, individual-specific taxonomic profiles have been demonstrated. Also for functional profiles, greater inter-individual than intra-individual variation is observable, at the metagenomic and metatranscriptomic level. Differences in functional profiles provide direct pointers to the functions involved in microbiome–host interactions will be quite novel and should be applied to the data to advance the field.

REVIEWER	Ana M Valdes University of Nottingham United Kingdom
REVIEW RETURNED	08-Oct-2018

GENERAL COMMENTS	This is a well thought clinical trial protocol testing the effect of FMT on adolescent obesity. A limitation that is not addressed is the transient effect of FMT seen with regards to metabolic syndrome in a recent RCT larger (n=38 Kootte et al 2017 PMID: 28978426) than the one cited in the manuscript (n=18). Although a strong effect is expected at 6 weeks, at least with regards to insulin sensitivity by week 18 the gut microbiome is expected to revert to its original state. This may also be the case for weight loss and it would be pertinent if the authors cited it and discussed it, not just the apparently more lasting effects of C. difficile treatment by FMT. The authors have stated that food diaries are part of the study. However, because fibre intake is expected to be an important contributor to any changes in metabolic syndrome and gut microbiome related changes it may be good to specifically address how this will be addressed in the data analysis phase.

REVIEWER	Leigh Greathouse Baylor University, USA
REVIEW RETURNED	18-Oct-2018

GENERAL COMMENTS	Overview: This is a well-designed study to assess the effect of FMT via encapsulated frozen stool on changes in BMI z scores among adolescents with obesity. The population samples size is adequately powered to test their hypothesis, and the assessment tools and measures are appropriate to the study. Some concerns and details need to be addressed to ensure completeness and awareness of possible pitfalls during collection and analysis of gut microbiome composition. Major concerns: 1. It is unclear why the authors are choosing a “homebrew” method of FMT preparation instead of going with a well-validated already prepared FMT capsule, like the one sold by Openbiome? I assume to save cost and allow gender matching? 2. The authors indicate they want to conduct 16S rRNA gene analysis to measure stool composition as a secondary measure. However, unless I missed it somewhere in the protocol, I see no mention of how, when, and where they will collect stool samples. This is a very crucial step and multiple papers have been written on how to accurately collect, save, store, and ship stool samples for analysis. Currently, if 16S rRNA analysis is the only downstream measure they wish to analyze, then the DNAgenotek stool collection kit with collection tube, spatula, toilet seat is the best option. I would assume at the very least you would collect stool at baseline and at 6 weeks; better would be to collect multiple time points during FMT to track composition with BMI changes, as well as, at the follow-up time points past six weeks to determine resilience of the FMT. These protocols are critical and need to be addressed as part of the study design. 3. In this same vein, the authors do not mention any methods regarding how they will extract DNA from stool, controls for

	contamination checking during DNA extraction, primers for sequencing, bioinformatics tools for taxonomic analysis, or statistics for analysis of the compositional data. Several papers have identified key steps during this process that can lead to spurious results, most prominently those from the Microbiome Quality Control Working Group (NCI). These protocols are critical and need to be addressed as part of the study design. Minor concerns: 1. Will you also look at appetite hormones such as GLP-1, CCK, leptin, ghrelin, PYY? Given they are critical in appetite regulation and weight control, and associated with the microbiome, I believe these would be important secondary endpoints. 2. How will you assess probiotics usage as criteria in exclusion/inclusion?

REVIEWER	Dave Gillespie Cardiff University, UK
REVIEW RETURNED	29-Oct-2018

GENERAL COMMENTS	This is a well written protocol describing a randomised controlled trial of gut microbiome transfer for the treatment of obesity in adolescents. I have some minor suggestions for revisions that may help improve the manuscript. INTRODUCTION • Other than the final sentence, I felt the final paragraph should have come a lot earlier in this section. Even as early as the first paragraph. I felt it set the scene nicely as was out of place towards the end of the section. METHODS AND ANALYSIS • I thought the rationale behind recipients only receiving gut microbiome from donors of the same sex required some reference to the literature where this is an established necessity. • The need for recipients receiving an equal number of capsules from each of the four same sex donors should be explained. • On Page 8 there was some inconsistency between 26-weeks and six-months. I think both were used interchangeably and it would be helpful to keep these consistent throughout. • On Page 11, I found the detail of the measurement of gut microbial composition rather light. Can more detail be provided? • Regarding the effect size this study is powered to detect – what does a 0.19 difference in BMI SDS mean, and is this effect meaningful? There was a lack of justification for this effect size, which I found concerning. • On Page 13, I would like some description of the following: o How sub-group analyses are going to be carried out o Which key secondary outcomes are going to be considered for sub-group analyses o More description of how you're going to do multiple imputation (how you're going to build an imputation model, how many imputations, etc.) o How you're going to carry out your per-protocol analysis (which violations will be considered, which analytical method/s, etc.) • I found the PPI section light. There was no real description of how PPI members input into the study design, for example. It sounds like beyond study design, there was no PPI involvement.

	• I found Table 3 unnecessary, as it was already sufficiently described in the text.

VERSION 1 – AUTHOR RESPONSE

REVIEWER 1 – Jed Friedman

We thank Dr Friedman for the time invested in the peer-review of our manuscript, and for the constructive feedback provided.

-Can you comment further on the primary outcome on which the study is powered?

Reply: The primary outcome is the BMI SDS for all the recipients at 6 weeks post treatment. Measuring the outcome via BMI SDS versus BMI has the advantage of allowing comparisons between different age groups and sex as the standard deviation scores are age and sex-independent.

On p. 12 it indicates a BMI difference of 0.19 at 6 wks - this seems small and would be subject to less clinically meaningful as say DEXA, OGTT, or a change in lipid levels.

Reply: The difference of 0.19 at 6 weeks reflects the BMI SDS difference rather than BMI difference. This will correlate to a 2-kg difference in weight approximately. We have since amended the respective statement in the Methods on sample size and power calculation, which now reads as:

"A study with 32 recipients per group will have 80% power at 5% significance level (two-sided) to detect a group difference of 0.19 in BMI SDS at 6 weeks after gut microbiome transfer, which is equivalent to a difference in weight of approximately 2 kg."

-Secondary outcomes beyond food questionaires including satiety measures might be important.

Reply: We agree that satiety is an important outcome to measure. We will indirectly assess this via quantity of food consumed throughout the trial using food questionnaires administered at baseline, 6 weeks, 12 weeks, and 26 weeks. In addition, as per our response to a comment by Dr Greathouse, we may also measure appetite hormones if the intervention is successful (these have not been budgeted for in our original design).

As the MB may only provide a small short-term effect on body weight unless diet is modified, this would seem to be more important to assess outcomes. SCFA are currently in vogue for insulin sensitivity and for metabolic effects and should be considered.

Reply: In this trial, our primary aim was to assess the efficacy of gut microbiome transfer as a mode of treatment for obesity in adolescents. As a result, we asked the recipients not to modify their diet throughout the trial. We agree with the reviewer that SCFA could be a useful indicator to assess. We are storing samples, and pending the results of the trial on primary and secondary outcomes, we may seek additional funding to measure SCFAs.

The central question beyond body weight are what determines whether this study can reveal important functional elements of the microbiome such as

(i) is the observed variability biologically meaningful?

(ii) is a measured microbial functional state identifiable?

Reply: We thank Prof. Friedman for this comment. We agree that revealing biologically significant variability in both the population structure and functional states of the post-transfer microbiome is

critical. This is particularly important given the commonly held assumptions about the dynamic nature of the human gut microbiota population.

We and others have demonstrated that a subset of the microbiome are conserved over the life-time of the host (Hum Microbiome J 2017;5-6:7-10), are impacted on by host genetics (Nat 2018;555:210-5 / Nat Genet 2016;48:1407-12), and yet remain malleable to gut microbiome transfer (Cell Metab 2017;26:611-6.e6). We contend that the integrative mixed 'omics' approach we are proposing will enable the identification of organismal (i.e. 16S amplicon and shot-gun metagenomics), metabolic (i.e. metatranscriptome and mass spectrometry), and phenotypic changes that occur following gut microbiome transfer. The utility of these comparisons is dependent upon the detailed phenotyping of this cohort, which provides the additional datasets that are an essential component of the meta-analysis to discover the biologically relevant associations. However, it is important to acknowledge that even in an intervention trial of this detail and complexity, the inter-relationships between the microbiome and host phenotype will at best remain associations until proven by subsequent studies.

Our pilot data support the ability of the proposed approach to identify variability in the microbiota. These data were obtained from 11 female recipients as part of the optimization process for this trial. The pilot data clearly identify a shift in the microbiome of the gut microbiome recipients (see attached confidential figure for reviewers and editor only). Moreover, the reproducibility is such that we have been able to identify specific organisms that change as a result of the gut microbiome transfer (data not shown).

For the metagenome, individual-specific taxonomic profiles have been demonstrated. Also for functional profiles, greater inter-individual than intra-individual variation is observable, at the metaagenomic and metatranscriptomic level. Differences in functional profiles provide direct pointers to the functions involved in microbiome–host interactions will be quite novel and should be applied to the data to advance the field.

Reply: We agree with the Prof. Friedman. We have clarified that we will include metagenomics analyses in Methods, with the addition of the paragraph below (sub-heading Gut microbiome composition, last paragraph):

"Metagenomic sequencing data will be analysed using default parameters of the HMP Unified Metabolic Analysis NEtwrok (HuMAnN2) (version 2; or later) after removal of short reads (minimum length 50 bases, trimmomatic version 0.33 or later) and human sequences using BMTagger. MaAsLin (version 0.0.4; or later) will be used to identify significant associations between microbial compositions, metabolomics data, and microbial functions."

REVIEWER 2 – Ana M Valdes

This is a well thought clinical trial protocol testing the effect of FMT on adolescent obesity. A limitation that is not addressed is the transient effect of FMT seen with regards to metabolic syndrome in a recent RCT larger (n=38 Kootte et al 2017 PMID: 28978426) than the one cited in the manuscript (n=18). Although a strong effect is expected at 6 weeks, at least with regards to insulin sensitivity by week 18 the gut microbiome is expected to revert to its original state. This may also be the case for weight loss and it would be pertinent if the authors cited it and discussed it, not just the apparently more lasting effects of C. difficile treatment by FMT.

Reply: We thank Dr Valdes for the very useful comment. We have consequently included the above-mentioned study and have expanded the respective section in the manuscript, which now reads as (Introduction, paragraph 4):

"Kootte et al. reported similar results at 6 weeks among 38 obese males (median age 56 years), but the improvements in both insulin sensitivity and gut microbiota composition reverted back to baseline at 18 weeks 23. Conversely, our group (unpublished data) demonstrated that gut microbiome composition in recipients changed after gut microbiome transfer to mimic the lean donor's gut microbiome, and that this effect was sustained 26 weeks after treatment. This indirectly indicates that it is possible to change the gut microbiome, using a healthy donor, with possible concurrent health benefits."

Of note, our unpublished data show that our protocol promotes microbiome changes that remain after 26 weeks (please see confidential figure). In addition, our primary outcome is BMI SDS and not insulin sensitivity (the latter is a secondary outcome).

The authors have stated that food diaries are part of the study. However, because fibre intake is expected to be an important contributor to any changes in metabolic syndrome and gut microbiome related changes it may be good to specifically address how this will be addressed in the data analysis phase.

Reply: We agree that fibre intake is likely to play an important role in gut microbiome composition, and consequently on associated metabolic effects. Our secondary analyses will include possible dietary effects (such as fibre intake), and this has since been clarified in our manuscript (Methods, sub-heading Statistical analyses, last paragraph):

"Our secondary analyses will include the examination of potential effects of diet (e.g. fibre intake) and physical activity levels on study outcomes."

REVIEWER 3 – Leigh Greathouse

This is a well-designed study to assess the effect of FMT via encapsulated frozen stool on changes in BMI z scores among adolescents with obesity. The population samples size is adequately powered to test their hypothesis, and the assessment tools and measures are appropriate to the study. Some concerns and details need to be addressed to ensure completeness and awareness of possible pitfalls during collection and analysis of gut microbiome composition.

Reply: The encouraging constructive comments provided by Dr Greathouse are very much appreciated, as is the valuable time invested in assessing our manuscript.

Major concerns:

1. It is unclear why the authors are choosing a "homebrew" method of FMT preparation instead of going with a well-validated already prepared FMT capsule, like the one sold by Openbiome? I assume to save cost and allow gender matching?

Reply: We thank Prof Greathouse for her comment. Our study includes the assessment of body composition using DXA scans as a critical component of our detailed inclusion criteria for donors, which is not accounted for by Openbiome. In addition, we are using a modification of Youngster et al.'s (JAMA 2014;312:1772-8) encapsulation method for the following reasons:

1) Regulatory restrictions associated with transferring capsules into New Zealand are such that we cannot use Openbiome, and there is also no local supplier;

2) We are matching for sex;

3) We contend that local dietary patterns, environmental toxin profiles, and seasonal variations may impact on the transfer if the donor material is sourced from another country;

4) Using our own material, we have total control over the time from encapsulation to treatment and validation of the cold-chain for the processed material; and

5) Cultural considerations may preclude the acceptance of transfer from unknown individuals from another country.

2. The authors indicate they want to conduct 16S rRNA gene analysis to measure stool composition as a secondary measure. However, unless I missed it somewhere in the protocol, I see no mention of how, when, and where they will collect stool samples. This is a very crucial step and multiple papers have been written on how to accurately collect, save, store, and ship stool samples for analysis. Currently, if 16S rRNA analysis is the only downstream measure they wish to analyze, then the DNAgenotek stool collection kit with collection tube, spatula, toilet seat is the best option. I would assume at the very least you would collect stool at baseline and at 6 weeks; better would be to collect multiple time points during FMT to track composition with BMI changes, as well as, at the follow-up time points past six weeks to determine resilience of the FMT. These protocols are critical and need to be addressed as part of the study design.

Reply: Stool composition will be analysed using 16S rRNA amplicon sequencing. However, we also want to retain samples for use in metagenomics and metabolomics.

Prof Greathouse is correct that 16S amplicon sequencing and metagenomics are extremely sensitive techniques for the characterisation of the microbial composition and function. There is no single method that guarantees that the DNA/RNA libraries generated will be 100% fully representative of the original sample. Freezing, refrigeration, or treatment with a stabilising agent all impact on the composition that is observed (PLoS One 2015;10:e0134802 / Sci Rep 2015;5:16350 / FEMS Microbiol Lett 2012;329:193–7). Therefore, standardization of the protocol within a study is essential.

We have standardized our sample collection protocol to enable immediate extraction. Briefly, the participant is given the bedpan liner (Onelink). They are asked to pass urine into the toilet prior to placing the tray on the toilet seat, pass the stools, cover the tray, and leave it in the bathroom, for immediate collection by a research team member. Using a small spatula, three different areas of the stool will be sampled; proximal, middle and distal and inserted into specimen containers (Onelink). The specimen containers will be immediately placed on ice and taken to the laboratory where they are frozen and stored at -80oC. DNA and RNA extraction is completed within 5 days of donation. Time to processing is recorded.

Sample collection will be performed at baseline prior to treatment and at 6 weeks, 12 weeks and 26 weeks post-treatment. Therefore, in total for each recipient, we will be collecting 4 stool samples which will provide important information on changes to the gut microbiome composition throughout the trial.

The detailed information above has since been added to our manuscript (Methods, sub-heading Gut microbiome composition, 1st paragraph).

3. In this same vein, the authors do not mention any methods regarding how they will extract DNA from stool, controls for contamination checking during DNA extraction, primers for sequencing, bioinformatics tools for taxonomic analysis, or statistics for analysis of the compositional data. Several papers have identified key steps during this process that can lead to spurious results, most prominently those from the Microbiome Quality Control Working Group (NCI). These protocols are critical and need to be addressed as part of the study design.

Reply: Prof Greathouse is correct and we apologize for the omission of this information from original manuscript. We have since added highly detailed information into our Methods section in the manuscript, under the sub-heading Gut microbiome composition:

"Sample collection will be performed at baseline prior to treatment and at 6 weeks, 12 weeks and 26 weeks post-treatment. Briefly, the participant will be given the bedpan liner (Onelink). They will be asked to: i) pass urine into the toilet prior to placing the tray on the toilet seat; ii) pass the stools; iii) cover the tray and leave it in the bathroom for immediate collection by a research team member. Using a small spatula, samples will be collected from three different areas of the stool (proximal, middle, and distal) and inserted into specimen containers (Onelink). The specimen containers will be immediately placed on ice and taken to the laboratory where they will be frozen and stored at -80oC. DNA and RNA extraction will be completed within 5 days of donation. Time to processing will be recorded.

All extractions will be performed using Qiagen-AllPrep DNA/RNA mini kit®, due to variation in extraction efficiencies with the different kits 59. However, once the DNA or RNA is extracted and archived, we will have a relatively stable record of the composition and activity of the flora.

Frozen faeces (~200 mg; weights will be recorded) will be subsampled from original faecal samples. All DNA and RNA isolations will be performed in a disinfected class II hood at room temperature. Briefly, stool samples will be incubated (10 min, room temperature) with vortexing (30 sec every 2 minutes) and treated with RLT Plus buffer (1.2mL; Qiagen) and 12µL beta-mercaptoethanol (Sigma-Aldrich). Acid-washed glass beads [1 ml; ≤106 µm (-140 U.S. sieve) (Sigma-Aldrich)] will be added to each sample and vortexed (10 min) on a TissueLyzer II (Qiagen). The supernatant will be removed and added to a QIAshredder spin column (Qiagen) and centrifuged (9000 rpm, 2 min, room temperature). The eluent will be added to an AllPrep DNA (Qiagen) spin column and centrifuged (30 sec, 14000 rpm, room temperature). The eluent and AllPrep DNA spin columns will be used for RNA and DNA extraction, respectively, according to the manufacturer's instructions. Finally, DNA and RNA will be eluted with EB buffer and RNase-free water, respectively, and aliquots stored at -80oC for downstream mixed omics analysis.

A series of blank samples (sterile saline) will be extracted in parallel to sample extractions to enable contamination testing. We will also extract ZymoBIOMICS™ Microbial Community Standard I (Even, Cellular Mix; Catalog #D6300) to determine potential bias in the extraction process.

For 16S amplicon sequencing, library preparation will be performed using an Illumina platform by a commercial provider (to be determined) using standard protocols for the SV3-4 region. Shotgun metagenomics sequencing will be performed by a commercial provider (to be determined).

All raw sequencing files will be cleaned to remove adaptors and primer sequences, and trimmed for sequence quality (Phred score<30).

Longitudinal analysis of gut microbiome data (i.e. change in alpha and beta diversity from baseline to 26 weeks in treatment and placebo group) will be performed on Qiime2 (version 2018.4 or later) using default parameters 60. PERMANOVA and Multivariate Association with Linear Models using MaAsLin (version 0.0.4; or later) 61 will be used to identify any significant differences in gut microbial communities and structure between treatment groups.

Metagenomic sequencing data will be analysed using default parameters of the HMP Unified Metabolic Analysis NEtwrok (HuMAnN2) (version 2; or later) 62 after removal of short reads (minimum length 50 bases, trimmomatic version 0.33 or later 63) and human sequences using BMTagger 64. MaAsLin (version 0.0.4; or later) 61 will be used to identify significant associations between microbial compositions, metabolomics data, and microbial functions."

Minor concerns:

1. Will you also look at appetite hormones such as GLP-1, CCK, leptin, ghrelin, PYY? Given they are critical in appetite regulation and weight control, and associated with the microbiome, I believe these would be important secondary endpoints.

Reply: Dr Greathouse makes a valid suggestion. However, appetite hormones are not part of the original design of the study because of their added cost. Nonetheless, if we can obtain additional funding and the intervention is successful we will definitely would consider assessing these hormones. But please note that we are measuring food intake as a clinical measure of appetite.

2. How will you assess probiotics usage as criteria in exclusion/inclusion?

Reply: We thank Dr Greathouse for pointing out our oversight, as probiotics intake should have been listed in Table 2 as an exclusion criterion. We have since amended the table accordingly, as participants were specifically asked about any prior consumption of probiotics before being recruited into the trial.

REVIEWER 4 – Dave Gillespie

This is a well written protocol describing a randomised controlled trial of gut microbiome transfer for the treatment of obesity in adolescents. I have some minor suggestions for revisions that may help improve the manuscript.

Reply: We are grateful to Dr Gillespie for the appraisal of our manuscript and the encouraging feedback provided.

INTRODUCTION

• Other than the final sentence, I felt the final paragraph should have come a lot earlier in this section. Even as early as the first paragraph. I felt it set the scene nicely as was out of place towards the end of the section.

Reply: Dr Gillespie makes a useful suggestion, and have since placed the respective paragraph as the first one of the Introduction.

METHODS AND ANALYSIS

• I thought the rationale behind recipients only receiving gut microbiome from donors of the same sex required some reference to the literature where this is an established necessity.

Reply: We primarily wanted to assess the efficacy of gut microbiome transfer as a modality of treatment for obesity and the result may potentially confounded by sexual dimorphism. Markle et al. 2013 (Science 2013;339:1084-88) demonstrated that caecal microbiota transfer from male non-obese diabetic mice (NOD) to female NOD mice prior to diabetes onset protected the female mice from pancreatic islet inflammation, autoantibody production and this was associated with increased testosterone in female mice. This study highlighted that there may be potentially sex-specific differences in the effect of gut microbiome on weight and metabolism. This clarification has since been described into the manuscript (Methods / Recruitment and eligibility criteria / Donors):

"We will recruit 8 donors (4 males and 4 females), as recipients will only receive gut microbiome from donors of the same sex. This is to enhance microbial variability and standardise the treatment via gut microbiome transfer. Treatment with gut microbiome from donors of the same sex will be done as there may be potentially sex-specific differences in the effect of gut microbiome on weight and metabolism as described by Markle et al. 33"

• The need for recipients receiving an equal number of capsules from each of the four same sex donors should be explained.

Reply: We have standardized treatment for recipients to reduce the variability that this could create. Thus, all recipients received the same number of capsules from the four same-sex donors. In addition, this also ensured that overall donor microbiome diversity was increased and delivered in reproducible fashion to all recipients. We have since added a comment to this regard into the manuscript (Methods, Study intervention, last sentence):

"All recipients will receive the same number of capsules from the four same-sex donors to standardized treatment and to ensure that overall donor microbiome diversity is increased and delivered in a reproducible fashion."

• On Page 8 there was some inconsistency between 26-weeks and six-months. I think both were used interchangeably and it would be helpful to keep these consistent throughout.

Reply: We apologise for this inconsistency and have since standardised the terminology to 26 weeks.

• On Page 11, I found the detail of the measurement of gut microbial composition rather light. Can more detail be provided?

Reply: We apologize for this oversight. We have since provided detailed description in the manuscript, as per our response to Reviewer 3's major comment #3.

• Regarding the effect size this study is powered to detect – what does a 0.19 difference in BMI SDS mean, and is this effect meaningful? There was a lack of justification for this effect size, which I found concerning.

Reply: The study needs to be powered to show the smallest meaningful change between groups as reflected in BMI SDS. We consider the 0.19 difference in BMI SDS (equivalent to approximately 2 kg difference in weight) to be a small but meaningful clinical change.

• On Page 13, I would like some description of the following:

- How sub-group analyses are going to be carried out

Reply: Pre-planned sub-group analyses will be carried out separately for males and females, since sex was the factor for stratification of participants during randomization.

-Which key secondary outcomes are going to be considered for sub-group analyses

Reply: We apologize for the oversight, as the word 'key' should not have been in that statement (this has since been rectified). Sex-specific analyses will be carried out for the same pre-specified secondary outcomes examined in the main trial analyses. We have since amended the respective sentence in the manuscript.

More description of how you’re going to do multiple imputation (how you’re going to build an imputation model, how many imputations, etc.)

Reply: Dr Gillespie makes a valid comment, as we agree that this information should have been included in the manuscript. A paragraph on this issue has since been added into the Methods section (sub-heading Statistical analyses, 2nd paragraph):

"Missing data on the primary outcome will be imputed using multiple imputations, which create multiple imputed datasets for the incomplete outcome variable that are analyzed using same regression models and combined for one inference. The Markov chain Monte Carlo (MCMC) method will be used to produce the parameter estimates, assuming the data are from a multivariate normal distribution and are missing at random. The SAS procedure, PROC MI, will be used which runs 200 iterations of the algorithm before selecting the first completed data set, and then allows 100 iterations

between each successive data set. The default minimum number of imputations is 5, and we plan to run 30 to allow for both within and between imputation variances."

How you're going to carry out your per-protocol analysis (which violations will be considered, which analytical method/s, etc.)

Reply: Information on this aspect is also provided in our revised manuscript (sub-heading Statistical analyses, 3rd paragraph):

"Per-protocol analyses will be carried out on those recipients without major protocol violations. A protocol deviation form will be used to record all major protocol deviations, and reviewed in a blinded fashion by the trial steering group prior to final data lock. The per-protocol population will be analysed using same regression models as the primary intention-to-treat (ITT) population to test the robustness of main trial findings."

I found the PPI section light. There was no real description of how PPI members input into the study design, for example. It sounds like beyond study design, there was no PPI involvement.

Reply: We apologize for the paucity of details in our initial submission. We have since expanded this section, which now reads as follows:

"Public input into the study design was provided in open meetings by the Northern A Health and Disability Ethics Committee, whose membership includes both clinical and lay persons, as well as Māori representatives (New Zealand indigenous people). Information on the trial was subsequently made available on social media platforms (e.g. Facebook), which allowed participants to read and contact the researchers if they wanted to participate. Participants were not involved in the development, recruitment of other participants, or conduct of the trial. All recipients will be asked about any possible adverse effects of treatment at specific time points throughout the trial; if any serious adverse effects are reported, a thorough follow-up will be conducted to investigate the incident. After completion of data analyses, all recipients will receive information about their individual results."

We have also amended the paragraph on dissemination of findings, which now reads as:

"Communication to the scientific community will be through high-profile international research meetings, as well as relevant national and regional meetings. We aim to publish findings in high-impact peer-reviewed international journals. Further, the research team will communicate the findings to the general public in New Zealand and overseas through our institution's Communications Manager. Relevant findings will be shared with the community in a culturally appropriate manner."

• I found Table 3 unnecessary, as it was already sufficiently described in the text.

Reply: We agree with Dr Gillespie that Table 3 does repeat some of the information provided in the text. However, we would prefer to keep this table in the manuscript, as it helps the reader visualize all individual assessments being carried out and their respective time-points. Nonetheless, if Dr Gillespie and the editor feels strongly about it, we would be prepared to remove Table 3 from the manuscript.

VERSION 2 – REVIEW

REVIEWER	Ana M Valdes School of Medicine University of Nottingham UK
REVIEW RETURNED	10-Dec-2018

GENERAL COMMENTS	The authors have adequately addressed all the reviewer comments

REVIEWER	Leigh Greathouse Baylor University, USA
REVIEW RETURNED	25-Dec-2018

GENERAL COMMENTS	The authors have done an excellent job in attending to the additions I suggested with regard to stool collection and microbiome analysis. I have one issue remaining with the stool collection. The investigators indicate, as I read it, they will have the participants pass stool while at the clinic and that clinic staff will be there to immediately collect stool. In my experience, expecting a participant to produce stool within a short time period is not possible, even within 24hrs. It is possible that by the time they show up to clinic, they may have already had their BM of the day, or possibly may not be able to have a BM that day. I would highly suggest you send them home with a stool collection kit (many are available commercially or you can make a kit yourselves to take home), and have them either bring it back with them or better yet include a box to mail the stool sample back to the clinic. This is the procedure used by the iHMP protocol. If the stool is collected in a tube with stabilization buffer and bead (OMNIgene gut kit), they are stable for up to 60 days and can be mailed in without worry of keeping cold. All of these details have been worked out and are available on the iHMP website. Separately, I would suggest adding a Food Frequency History diet analysis to your baseline assessment. Though your current dietary analysis study design is adequate, adding an FFQ would improve it even further, and allow your team to assess FMT uptake as a function of prior dietary history (https://dietassessmentprimer.cancer.gov/approach/table.html). This may be important in understanding response to and resilience of the FMT.

REVIEWER	David Gillespie Cardiff University, UK
REVIEW RETURNED	19-Dec-2018

GENERAL COMMENTS	I was satisfied to responses for most of the comments I provided. I do however have a couple of final points that should be considered before I can recommend publication: • The authors propose to conduct sex-specific analysis. As this is a subgroup within their trial and they're interested in effect modification, I think the correct analysis would be extending their primary analysis by including an interaction term between sex and trial arm (see Wang, R., Lagakos, S.W., Ware, J.H., Hunter, D.J. and Drazen, J.M., 2007. Statistics in medicine—reporting of subgroup analyses in clinical trials. New England Journal of Medicine, 357(21), pp.2189-2194.)

	• The authors' current proposal for a per-protocol analysis may be prone to considerable selection bias, and they may want to consider using instrumental variables methods to adjust for protocol violations, particularly if they are primarily departures from randomised treatment (see White, I.R., 2005. Uses and limitations of randomization-based efficacy estimators. Statistical methods in medical research, 14(4), p.327.)

VERSION 2 – AUTHOR RESPONSE

REVIEWER 2 – Ana M Valdes

The authors have adequately addressed all the reviewer comments.

Reply: We thank Dr Valdes once again for the time invested in the peer-review of our manuscript, and we are glad to hear that her comments have been adequately addressed.

REVIEWER 3 – Leigh Greathouse

The authors have done an excellent job in attending to the additions I suggested with regard to stool collection and microbiome analysis. I have one issue remaining with the stool collection. The investigators indicate, as I read it, they will have the participants pass stool while at the clinic and that clinic staff will be there to immediately collect stool. In my experience, expecting a participant to produce stool within a short time period is not possible, even within 24hrs. It is possible that by the time they show up to clinic, they may have already had their BM of the day, or possibly may not be able to have a BM that day. I would highly suggest you send them home with a stool collection kit (many are available commercially or you can make a kit yourselves to take home), and have them either bring it back with them or better yet include a box to mail the stool sample back to the clinic. This is the procedure used by the iHMP protocol. If the stool is collected in a tube with stabilization buffer and bead (OMNIgene gut kit), they are stable for up to 60 days and can be mailed in without worry of keeping cold. All of these details have been worked out and are available on the iHMP website.

Reply: We thank Dr Greathouse for her positive feedback. We agree with you that some participants may experience difficulty in producing stool sample during their visits to the clinic for their assessments. We advise them to try not to have a bowel movement in the morning prior to their visit, having it in the clinic instead. For those participants who are unable to produce a stool sample during their visit, they are provided with a stool collection kit to take home and detailed instructions on how to collect the stool sample. This kit is made up of: i) instructions on how to use the stool collection kit; ii) specimen container; and iii) bedpan liner. Once the stool has been collected in the home environment, the specimen container is immediately placed into their home freezer, and it is kept there until it is delivered to the research team. We have since added the above information into a new paragraph in the Methods, under the subheading "Gut Microbial Composition".

Separately, I would suggest adding a Food Frequency History diet analysis to your baseline assessment. Though your current dietary analysis study design is adequate, adding an FFQ would improve it even further, and allow your team to assess FMT uptake as a function of prior dietary history (https://dietassessmentprimer.cancer.gov/approach/table.html). This may be important in understanding response to and resilience of the FMT.

Reply: We agree with Dr Greathouse that this information is useful. We do have food frequency questionnaires as part of our study protocol, which are administered to the participants at baseline, and at 6, 12, and 26 weeks post treatment. The Reviewer might have accidentally missed this information, which is provided in the second paragraph under the subheading "Dietary intake" in our manuscript:

"The New Zealand Adolescent Food Frequency Questionnaire (NZAFFQ)53 will be administered at baseline and weeks 6, 12, and 26. The NZAFFQ was developed for and validated in New Zealand adolescents aged 14 to 18 years53."

REVIEWER 4 – Dave Gillespie

I was satisfied to responses for most of the comments I provided. I do however have a couple of final points that should be considered before I can recommend publication:

• The authors propose to conduct sex-specific analysis. As this is a subgroup within their trial and they're interested in effect modification, I think the correct analysis would be extending their primary analysis by including an interaction term between sex and trial arm (see Wang, R., Lagakos, S.W., Ware, J.H., Hunter, D.J. and Drazen, J.M., 2007. Statistics in medicine—reporting of subgroup analyses in clinical trials. New England Journal of Medicine, 357(21), pp.2189-2194.)

Reply: We agree with Dr Gillespie's comment and have added more details on subgroup analyses in the Methods ("Statistical analyses", paragraph 1), so that it reads as:

"Planned subgroup analysis by sex will be conducted on primary and secondary outcomes to evaluate the consistency of main treatment effects in males and females, by including an interaction term between sex and treatment group in the main model. If a significant interaction effect is found, separately subgroup analyses will be conducted to estimate the treatment effects in specific subgroups."

• The authors' current proposal for a per-protocol analysis may be prone to considerable selection bias, and they may want to consider using instrumental variables methods to adjust for protocol violations, particularly if they are primarily departures from randomised treatment (see White, I.R., 2005. Uses and limitations of randomization-based efficacy estimators. Statistical methods in medical research, 14(4), p.327.)

Reply: Dr Gillespie makes a valid comment. However, it is important to point out that it is not possible for participants to violate the treatment protocol since there is a single course of FMT at the start of the study. The potential protocol violations of relevance in our study would be radical changes in diet and lifestyle, such as moving from sedentary habits to marathon training. Thus, we do not believe selection bias will be an issue, especially since any cases of protocol violations as per the research team's assessment in a blinded fashion will be scrutinized by the data monitoring committee, before final data lock and analysis.