Reviewer Response

Editor comments

Additional Editor Comments:

I concur with the reviewers that the concept of the manuscript is relevant and timely, but that a more balanced perspective should be provided. As stated by one of the reviewer, the recent fatalities linked to FMT would be the logical examples to tone down the hypotheses and conceptual framework to include more about safety.

- We agree with the editor and reviewer 1 that we did not sufficiently emphasize safety, and have edited our manuscript accordingly. Specifically, we now explicitly discuss that the rational donor selection should be performed on a set of donors who have already been screened for a variety of safety considerations, and that clinicians should never compromise on safety. We expand more on this point in our response to Reviewer #2.

Reviewer #1

This is an interesting paper that focuses on using biology-based hypotheses to select donors for FMT trials. Although the manuscript is in general well-written, there are some concerns.

The idea of using different case studies to test the different potential mechanisms of FMT is ok, but I am surprised that only the most favorable hypothesis is tested in each of the case studies separately. It would make more sense to test all four of them for the different cases and see if the selection based on the most likely biological mechanism would be most promising.

- It seems like the reviewer is suggesting that FMT trials should be used to distinguish between the four potential biological mechanisms. While this may be possible in theory, in practice most trials will likely not be powered to perform these analyses, since you would need a large number of patients in the FMT arm in order to test all four hypotheses.
- In contrast, our framework attempts to address the practical question of how to maximize the probability of a successful clinical trial according to its primary outcome: clinical efficacy. In the case where a clinician has a strong *a priori* hypothesis for the biological mechanism(s) underlying the disease, this means selecting donors based off of that most favorable hypothesis.
- We agree that in some cases, clinicians may not actually know which mechanism is most favorable, or may have multiple competing mechanistic hypotheses which are equally likely to be involved. In this case, we suggest that donors should be cycled and paired with an adaptive trial design in order to maximize the probability of clinical trial success. The reviewer is correct that cycling through donors strategically may allow clinicians to test which mechanistic hypothesis may be involved by selecting donors

through a combination of strategies and seeing whether one strategy led to greater FMT success than others.
- We have added a note clarifying these points in the “Cycling healthy donors in adaptive trials” section.

I am surprised that the study of Fuentes et al 2017 in ISMEJ is not discussed et all as this FMT study included the majority of the proposed mechanisms that are described in this manuscript, including butyrate production (capacity), signature microbes for health/disease, and donor-recipient matching.

- While Fuentes et al did not perform rational donor selection, their paper is an excellent example of many retrospective analyses that we discuss, illustrating both the potential for and limitations of performing retrospective studies on factors determining FMT success. We have included this citation in our discussion of retrospective analyses (section “Discovery-based retrospective analyses”).

I also lack a critical view on the described approach in this study in the discussion section. For example, it does not discuss survivability of microbes during the FMT procedures which are generally done under oxic conditions. Also the mode of delivery is not discussed. One can imagine that for CD infusion into the small intestine might be more effective in case the inflammation occurs in the ileum. Also, the authors focus mainly on DNA based approaches when studying the butyrate producers while it is evident that butyrate production is dependent on the metabolism of the organism. Placebo effect is also not discussed while it has been published that this occurs. Last but not least, the authors do not discuss that feces is not synonymous for intestine and that looking at fecal microbiome alone can be limiting.

- The reviewer is correct that there are many logistical and scientific considerations involved in performing FMT trials which we did not address in this paper. However, many of these (such as delivery mode, placebo effect, and organism survival during FMT prep and procedure) are outside the scope of this paper, which presents a framework to select which donors will be used but does not touch on these other aspects of designing and performing an FMT trial. Other researchers have attempted to address these questions, and they continue to remain open questions in the field actively being researched.
- We agree with the reviewer regarding DNA-based approaches, and think that directly measuring community functionality will likely be a better measure of community function. That said, we also recognize the benefit of DNA-based approaches if resources to perform functional assays or metagenomics are not available, which is often the case in limited exploratory clinical trials.
- We also agree that fecal-based assays may miss out on important parts of the GI physiology (e.g. mucus), which may be difficult to address through FMT generally. We have added a note about regional variation in the microbiome in the discussion as a limitation of using the fecal microbiome as a representative of the entire GI physiology.

- We have added a limitations paragraph to the discussion, which touches on these points.

Reviewer #2

This submission proposes a conceptual framework of rational donor selection in FMT clinical trials. The authors propose that clinical FMT trial failures, for example in IBD, may relate to suboptimal donor selection and propose a rational design approach to identify optimal donor microbiota communities and inferred metabolic capacity to facilitate clinical trial efficacy. Four theoretical models of why FMT may fail in patients are proposed, with some metagenomics and metabolomics data presented to support the basis for the conceptual framework approach. On the whole, the study represents an interesting hypothetical overview aimed at guiding the physician/scientist when considering clinical FMT trial design, especially in patients where standardized microbiota therapy fails. The manuscript also uses simulated microbiota community models to demonstrate that retrospective FMT analysis is generally underpowered to instruct on candidate therapeutic microbial communities, which seems somewhat counterintuitive in the context of the conceptual framework models that target specific beneficial or pathogenic taxa.

- To clarify, in this paper we touch on two different questions: (1) how do you design an FMT trial to maximize the primary clinical outcome? and (2) given a completed trial, how do you retrospectively analyze the results to identify taxa mediating FMT response? In this manuscript, we almost exclusively focus on question #1: clinicians who have strong hypotheses about which model is most likely to be involved should use their *a priori* hypothesis to drive donor selection. This *a priori* hypothesis is derived holistically, combining information from published cross-sectional studies, mechanistic investigations, completed FMT trials, and their own clinical experience. We touch on question #2 in the power simulation, showing that most FMT trials are not powered to identify the taxa mediating FMT response. This leads to the counterintuitive aspect the reviewer mentioned, where a clinician can design a clinical trial based on their *a priori* holistically-derived hypothesis that individual taxa are important to FMT response, but then not be powered to identify those taxa in their retrospective analysis.

The manuscript would benefit from consideration of the following points:

1. Since this manuscript was submitted a serious adverse event involving donor transfer of antibiotic-resistant bacteria resulting in serious disease and fatality in two immunodeficient FMT recipients. The FDA has now implemented a number of new restrictions and safety measures, as well as reiterating the investigational nature of FMT trials in patients. As written this experimental therapy is described as standard-of-care by the authors. Appropriate revision is required, especially in light that the FDA has expressed concerns about sourcing universal fecal banks for treatment. This change in practice and new guidelines implemented needs to be adequately addressed.

- We have now explicitly included a paragraph on the paramount importance of selecting donors who are healthy through elimination of donor stool that may transmit known or possible pathogens, and that this requirement supersedes any other consideration regarding donor selection. In this new section, we have highlighted the recent patient deaths as well as the FDA's issuance of new screening criteria aimed at eliminating donors with MDRO carriage. We have also reiterated that the treatment is investigational and has not been approved by the FDA for any indication. While an unusual situation, it is indeed the case that this investigational treatment has been defined in clinical guidelines as standard of care for multiply recurrent CDI. Our comment in this manuscript on FMT as standard of care refers narrowly to its use in the prevention of multiply recurrent CDI.
- The new FDA guidelines now recommend screening for MDROs, but don't say anything in particular about universal fecal banks. In fact, sourcing donor samples from universal stool banks should actually be a safer approach because stool banks can update their screening guidelines and apply them across-the-board to all of their stool samples in all of the studies they're involved in. This theory is supported by data from OpenBiome, the stool bank we are affiliated with, which has performed nearly 50,000 FMTs with no reported serious adverse events definitively linked to the FMT material.
- FDA has also not expressed concerns regarding the use of universal stool banks in the wake of the event, or none of which we are aware. Rather, FDA's public comments regarding stool banking are instead focused on identifying a suitable regulatory mechanism for their activities. Since issuing its guidance of enforcement discretion (in essence stating that it will not enforce the need for an IND to treat patients with multiply recurrent CDI), the FDA has sought public comment on new guidances that would ensure that the operations of universal stool banks receive regulatory oversight -- a necessary and appropriate adjustment.
- However, discussing the appropriate screening criteria for FMT donors and the relative merits of stool banks vs. directly sourced donors is out of the scope of this paper. In our paper, we intended to present stool banks as a good resource for rational donor selection because they have a wide variety of screened donors available for FMT studies. Implicit in this is that all donors have undergone rigorous screening procedures to ensure that they are safe for FMT, an assumption we have now made explicit throughout the revised paper.

2. In light of the risks now proven to be associated with FMT, the perspective should be toned down and better balanced. The proposal that the physician/scientist should select a donor from a fecal menu based on inferred metagenomics and a hypothetical disease mechanism – although interesting - seems somewhat irresponsible in this regard. This point is highlighted by our poor understanding of exactly what disease mechanisms and beneficial microbial communities are actually protective in recurrent C. difficile infection, arguably the only disease where FMT efficacy is clearly demonstrated.

- We agree that we did not emphasize our assumption that donors have already met certain health screening criteria explicitly enough. Implicit in our discussion was that first and foremost, anyone who might be compromised for safety is excluded as a donor. Then, donors can be rationally selected from among this pool of screened donors. Even after being screened, we expect that healthy microbiomes will still have a range of phenotypes. This is the range and variability from which we are suggesting physician/scientists rationally select their donors. We have explicitly clarified that donors should be selected from a pool of carefully screened donors in the manuscript to clarify this point.
- We agree that our poor mechanistic understanding of FMT success is a current limitation in the field, but disagree that this is an argument against performing rational donor selection in FMT studies. In fact, because almost all disease models lead to a donor selection strategy that involves giving patients stool from one healthy donor, the risk is thus the same no matter what biological mechanism is at play. In the one case where we suggest pooling multiple donors to increase the chance of finding rare taxa, risk from FMT might be increased and we have now explicitly clarified this in the text.
- Beyond these considerations, it is important to note that performing rational donor selection has an important side effect of increasing overall patient safety by improving the efficiency of clinical trials. The aim of performing rational donor selection is to maximize the power to detect clinical efficacy, which allows clinicians to get the scientific understanding they need while testing on the smallest number of human subjects. By performing rational donor selection, clinicians are actually also increasing the safety of the study because they will need fewer patients to discover a clinical effect.
- Finally, we have also clarified that we do not present a framework for deciding *whether* or not an FMT trial should be pursued, and such a framework is out of the scope of this paper. We have added a sentence clarifying this intention.

3. The above point is highlighted further by the widely recognized problems associated with predicting microbial function based on metagenomics data, especially 16S rDNA community profiling. This is evident in Fig. 2 which demonstrates a poor correlation between butyrate producers stratified by FMT efficacy in three reported IBD trials. This genus-level analysis is problematic since it makes the assumption that all taxa are conserved SCFA producers. More importantly, the most significant and abundant butyrate producers (the Eubacteria taxa) are excluded from the analysis.

- We agree that the analysis presented in Fig. 2 could be greatly improved with data at a finer taxonomic resolution. However, our intention in this paper was to present examples which would be reasonably accessible to clinicians, and as such we used a simple heuristic for assigning butyrate-producing taxa which has previously been published.
- Our case study in Figure 2 is purely an example for how a clinician might approach this problem, and is not intended to be anything more than a simple illustration. We have edited the text to clarify this intention. Despite this case study's limitations, we believe that it actually strengthens the manuscript considerably more than a toy example with

cleaner results but less applicability. By tackling a real-world example with real data, we demonstrate the process of rational donor selection and also highlight the difficulties involved in doing this work in practice.

- More specifically addressing the reviewer's point about Eubacteria: we had originally removed the Eubacteria taxa because the butyrate-producing taxa identified in Vital et al. 2017 (E. hallii, E. rectale, and E. ventriosum) were reported at the species-level, and do not comprise one genus with conserved butyrate production. We attempted to re-do this analysis with the excluded taxa included, but were unable to find any of these three Eubacteria reported in Vital et al. in our data. Specifically, our Green Genes-assigned ASV table contained only the following taxa:
    - Eubacterium genus-level taxa: g__[Eubacterium];s__biforme, g__[Eubacterium];s__dolichum, g__[Eubacterium];s__cylindroides, and g__[Eubacterium];s__
    - We also looked for ASVs assigned as Lachnospiracea_incertae_sedis at the genus-level, since this was the genus-level assignment for E. hallii, E. rectale, and E. ventriosum provided in Table S2 of Vital et al. We did not find any species of this genus in our data.
    - We have made a note about this limitation in the IBD case study, and added supplementary figures showing the abundance of each genus in each donor sample (Supp Figs 1-3).

4. Because a genus-based analysis is also used for the simulated models in Fig. 4, it appears to further underestimate the power required to retrospectively mine specific disease-associated and protective taxa that form part of a conceptual framework. Since strain-level biology is now also implicated in disease pathogenesis, e.g. in obesity, this does raise the question whether the metagenomics approach is optimal in the rational design of microbial therapy. The authors consider metabolic pathways but could expand on this discussion point, especially in light that there is much functional redundancy even in dysbiotic communities.

- We agree that genus-level retrospective analyses are suboptimal and unable to discover strain-level associations. However, as in the IBD case study in Fig 2, this analysis was intended to be primarily illustrative, and a full discussion of how to approach rational design of microbial therapies is out of the scope of this paper.
- We agree that the genus-level analysis underestimates the power needed to identify OTU, ASV or strain-level associations, given that the vast majority of OTUs/ASVs are not differentially abundant. However, given that even this conservative analysis fails to yield useful results with a typical trial size, it follows that FMT trials would be even less powered to identify associations at OTU/ASV/strain-level resolution. We have added this to the discussion of the results.
- We have also added further discussion on the limitations of taxonomy-based analyses vs functional data in the limitations paragraph in the discussion.

5. The conceptual framework described does not adequately consider mucosal-associated and regional (spatiotemporal) differences in microbiota communities, both relevant for example in Crohn's disease.

- The reviewer brings up an interesting point that different locations of the GI tract have different microbial community compositions. These differences could be considered in the selection process, for example by selecting donors with a high abundance of mucosal-associated taxa or by ensuring that functional assays mimic the environment where their mechanism of action is expected to take place (i.e. the mucus layer). We have added notes about considering mucosal-associated bacteria in the limitations discussion section.

6. The methods would benefit from more experimental detail, as would the figure legends in order to better appreciate the content and interpretations made.

- We apologize for the scarcity of experimental detail, and have expanded the methods section and figure legends accordingly. Thank you for this constructive feedback.
- We also note that the metabolomics data used in the "community functionality" case study was generated as part of a paper which has now been accepted for publication but is not yet published. We have cited this paper in our manuscript and updated the data availability in the methods.