OXFORD

Original Article

Re-Evaluating the Evidence for Faecal Microbiota Transplantation 'Super-Donors' in Inflammatory Bowel Disease

Scott W. Olesen[a,b] Ylaine Gerardin[c]

[a]OpenBiome, Cambridge, MA, USA [b]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA [c]Finch Therapeutics, Somerville, MA, USA

Corresponding author: Scott Olesen, 2067 Massachusetts Avenue, Cambridge, MA 02140, USA. Email: solesen@openbiome.org.

Abstract

Background: Faecal microbiota transplantation [FMT] is a recommended treatment for recurrent *Clostridioides difficile* infection, and there is promise that FMT may be effective for conditions such as inflammatory bowel disease [IBD]. Previous FMT clinical trials have considered the possibility of a 'donor effect', that is, that FMT material from different donors has different clinical efficacies.
Methods: Here we re-evaluate evidence for donor effects in published FMT clinical trials for IBD.
Results: In ten of 12 published studies, no statistically significant donor effect was detected when rigorously re-evaluating the original analyses. One study showed statistically significant separation of microbiota composition of pools of donor stool when stratified by patient outcome. One study reported a significant effect but did not have underlying data available for re-evaluation. When quantifying the uncertainty on the magnitude of the donor effect, confidence intervals were large, including both zero donor effects and very substantial donor effects.
Conclusion: Although we found very little evidence for donor effects, the existing data cannot rule out the possibility that donor effects are clinically important. Large clinical trials prospectively designed to detect donor effects are probably needed to determine if donor effects are clinically relevant for IBD.

Key Words: Faecal microbiota transplantation; inflammatory bowel disease; donor effect; super donor

1. Introduction

The human microbiome is increasingly understood to play a key role in health and disease.[1] Faecal microbiota transplantation [FMT], the infusion of a healthy person's stool into a patient, is one method for manipulating the gut microbiome.[2] FMT is recommended for treatment of recurrent *Clostridioides difficile* infection.[3,4] Although our understanding of the specific mechanisms by which FMT cures *C. difficile* infection is still developing,[5,6] FMT is being investigated as a therapy for dozens of other microbiome-related indications.[2,7,8]

A key challenge in identifying FMT's specific mechanism, or mechanisms, is the complexity and diversity of human stool. Stool is a mixture of bacteria, viruses, fungi, microbe-derived molecules and host-derived molecules that varies enormously from person to person.[9,10] It has therefore been hypothesized that different stool donors, different FMT material, or different matches of donors and recipients could have different abilities to treat disease. This concept has been referred to with terms such as 'donor effect', 'super-donor' and 'super-stool'.[8,11–14]

If FMT's efficacy varies widely across stools or donors, then rational selection of FMT material based on biomarkers predictive of efficacy could improve the clinical practice of FMT.[8,13,15–18] Differences in FMT efficacy between stool donors and specific stools could be also an important starting point for scientific investigations into the 'active ingredient' in FMT donations.[13] It is therefore important to quantify the extent of donor variability in indications where FMT is a promising treatment.

1

Although there is little evidence of a donor effect in the context of *C. difficile* infection,[19–21] the condition for which the use of FMT is best studied, multiple studies using FMT to treat inflammatory bowel disease [IBD] have tested for donor effects,[2,11,12,17,22–29] and some have reported statistically significant results. However, there are reasons to be sceptical of the clinical implications of these intriguing findings.

First, previous reports[15,30] found that, in simulations of clinical trials, studies with numbers of patients similar to the number used in extant studies would be unlikely to discover a donor effect unless they were very large. Statistically significant results from underpowered studies may represent discoveries of real effects, but they may also represent false positives.

Second, in all previous studies, identifying a donor effect was a *post hoc* analysis, and in no case was multiple hypothesis correction, a critical methodology in *post hoc* analyses, employed. In addition, we show herein that, in at least two previous reports, the specifics of the clinical design require adjustments to the statistical tests used to avoid inflating the apparent evidence for a donor effect.

Third, previous reports mostly analysed donor effects by testing null hypotheses that are biologically unlikely: either that all donors produce stool with exactly equal clinical efficacy, or that particular features of stool, such as its bacterial community diversity, are completely unrelated to clinical efficacy. However, it is highly improbable that all stool has precisely the same therapeutic effect: if a study is large enough, the difference between two donors' treatment efficacies would almost certainly become statistically significant. We therefore propose that, rather than merely asking if donor effects exist, we should ask about the magnitude of the donor effect and whether it is clinically relevant. In other words, rather than merely testing for the statistical significance of donor effects, we should also assess effect sizes.[31]

Here we re-evaluate the existing evidence for donor effects and discuss the implications of that evidence for clinical trial design and clinical practice. First, we establish an ontology of the various concepts referred to as 'donor effects'. Second, we lay out a rigorous framework for identifying and quantifying donor effects. Third, we re-evaluate the existing literature using the conceptual ontology and the rigorous framework. Finally, we discuss the implications of this re-evaluation for future FMT research in IBD.

2. Methods

2.1. Distinguishing types of stool superiority

Discussions of 'super-donors' and 'super-stool' have suggested that stool might be superior in at least four distinct but conceptually related senses. To avoid confusion in our re-evaluation of the evidence for donor effects, we distinguish between these definitions of stool superiority:

1. *Donor superiority*, in which particular donors are associated with better clinical outcomes for the recipient patients. For example, Moayyedi et al.[22] tested whether a particular donor ['donor B'] was associated with better patient outcomes, compared to the other donors.
2. *Donor characteristic superiority*, in which donors with particular characteristics are associated with better outcomes. For example, donor age, diet and host genetics have been suggested as potential factors in donor superiority.[12,28,32]
3. *Material characteristic superiority*, in which stool donations, pools of stool or other faecal microbiota preparations that have some particular characteristic are associated with better outcomes. For example, Vermeire et al.[33] tested whether stools with higher bacterial diversity are associated with better outcomes.
4. *Donor–recipient match superiority*, in which certain combinations of donors and recipients are associated with better outcomes,[11,12,17,18,25,34–36] analogous to how donors' and recipients' blood types are matched for blood transfusions. For example, FMT studies have tested whether stool from 'related' donors,[37] typically defined as first-degree relatives but also sometimes including spouses or partners, is associated with better or worse patient outcomes.

These types of superiority are conceptually related and not mutually exclusive. For example, imagine that female donors were associated with better patient outcomes than male donors, consistent with 'donor characteristic superiority' [definition 2]. However, because donors have a clinical effect on the patient only via their donation, the sex of the donor cannot be the molecular mechanism by which some donations are more efficacious than others. It would have to be that sex determined, or at least correlated with, some component of the stool that made those donations more effective. In other words, there must be an underlying 'material characteristic superiority' [definition 3] that correlates with donor sex. By a similar argument, even if donor–recipient match superiority is the most accurate model, simpler types of superiority may be more parsimonious.[14]

2.2. Identifying and quantifying donor effects

We used a two-step framework for characterizing donor effects. First, we used statistical tests to characterize the strength of the evidence for the reported type of donor effect. Statistics were computed using the R programming language[38] [version 3.6.0]. To test the hypothesis that there is any difference in efficacy between donors [or pools of donors], we used the Fisher–Freeman–Halton test [function *fisher.test*] on $2 \times N$ contingency tables, where N is the number of donors [or pools]. For 2×2 contingency tables [e.g. to test if one donor has a different treatment efficacy than all the others], we used Fisher's exact test [function *exact2x2*][39] and the mid-p value to prevent the tests from being overly conservative.[40] To test for material [i.e. donation or pool] characteristic superiority, we used the Mann–Whitney U test [*wilcox.test*] or Kruskal–Wallis test [*kruskal.test*]. When 16S rRNA gut microbiota composition data were available [see below], we tested for separation in donor microbiota compositions using PERMANOVA[41] [function *adonis* in the R *vegan* package].[42]

Second, whether or not an effect was detected, we use a random effect logistic regression [function *glmer* in the R *lme4* package][43] to quantify the effect size. The regression model assumes that the log odds of the efficacy of different donors [or pools of stool] is Gaussian distributed. The regression returns the standard deviation of that distribution, which is a measure of the 'spread' in donors' efficacies. For clarity, we convert the regression's estimates into the absolute difference in efficacy rates between a 90th percentile donor [a very high-quality donor] and a 10th percentile donor [a poor donor]. For example, an effect size of 30 percentage points means that, if a low-quality donor is 10% efficacious, then a high-quality donor would be 40% efficacious.

2.3. Studies re-analysed

In the re-analysis, we included interventional FMT clinical studies that tested the hypothesis that one donor or donation characteristic was associated with improved IBD patient outcomes. A two-step process was used to identify studies. First, citations

from an existing systematic review,[25] narrative review[12] and expert opinion[8] were evaluated for FMT IBD studies that searched for a donor effect or published sufficient data to allow for re-analysis. Second, a PubMed search was designed to include all the studies identified in the first step and to gather similar studies. The search was for MeSH terms 'fecal/faecal microbiota transplantation' or 'biological therapy', and MeSH terms 'inflammatory bowel disease' or 'ulcerative colitis' or 'Crohn disease', and keyword 'donor'. The search identified 161 publications, of which 14 were included in this study [two of those publications[17,44] were follow-on studies, so we report on 12 clinical studies]. Of the remaining publications, 80 were about therapies other than FMT, 45 were not human clinical studies, eight did not have sufficient data to test for donor effects [because of small numbers of patients or donors, insufficient contrast in patient outcomes, or a lack of microbiome samples], seven did not make the underlying data available for re-analysis, five were *C. difficile* studies in which IBD was not the primary outcome, and two were not available in English [Supplementary Table 1].

For each included study, we considered only a single dichotomous outcome, whether a patient achieved the trial's primary clinical end point. When 16S rRNA gut microbiota composition data were available, we tested for associations between donor material α-diversity and patient outcomes,[18,27,28,33,35] because part of the motivation for using pools of stool from multiple donors is the hypothesis that high bacterial diversity in FMT material is beneficial.[15,23,34,45] When possible, we also searched for separation in the microbiota compositions of donors according to the outcomes of their associated patients, following the example of Jacob et al.[24]

2.4. Microbiome data preparation
We re-analysed raw 16S rRNA gene sequencing data from four studies [Kump et al., study accession PRJEB11841; Jacob et al., accession PRJNA388210; Goyal et al., accession PRJNA380944; Nusbaum et al., accession PRJNA438164] using the programs QIIME 2[46] [version 2019.10] and Deblur.[47] Paired reads were joined [*vsearch* plugin, default parameters], quality filtered [*quality-filter* plugin, default parameters] and denoised using Deblur [trim length 253 nucleotides, 1 minimum reads, otherwise default parameters]. For Goyal et al., the sequencing data were single-ended [no joining required] and the trim length was 150 nucleotides. Alpha-diversity was computed by down-sampling all samples to the minimum number of denoised read counts across samples and then using the Shannon metric [*diversity* plugin]. When a donor had multiple associated samples, we associated a single α-diversity with each donor by computing the mean α-diversity over that donor's samples. Beta-diversity, used for the PERMANOVA tests, was computed with the Bray–Curtis metric. When a donor had multiple samples, we used only the first sample from each donor.

2.5. Code and data availability
Computer code and underlying data to reproduce the results are available online [DOI: 10.5281/zenodo.3780184, https://www.github.com/openbiome/donor-effects].

3. Results
The results of the re-evaluation of the evidence for a donor effect in IBD from a selection of FMT clinical studies are summarized in Table 1.

3.1. Rossen et al. 2015
In this study,[11] 23 ulcerative colitis patients were treated with FMT. Seventeen patients received stool from one of 15 donors; six patients received stool from two different donors. The outcome was clinical remission and endoscopic response at 12 weeks. In the original publication, the investigators reported on the performance of three donors [Table 2]. There was no evidence of a donor effect [mid-$p > 0.05$, Fisher's exact test] for any of the three donors. However, the study is small enough that the data are not inconsistent with large variability between donors. For example, the upper 95% confidence interval on donor A's odds ratio of a successful patient outcome compared to all other donors is 29. In other words, the data do not provide evidence of a donor effect, but they also cannot definitively rule out a strong donor effect.

A follow-up publication[17] reported an association between one bacterial taxon [*Ruminococcus gnavus*] in donors and patient outcomes, but this result was not statistically significant, even before multiple hypothesis correction [$p = 0.08$].

3.2. Moayyedi et al. 2015
In this study,[22] 38 ulcerative colitis patients were treated with FMT. Each patient was treated with stool from one of six donors. The outcome was remission at 7 weeks. Donors were used according to an adaptive process, described in the original publication. Briefly, at the start of the trial, patients were treated with stool from one of two donors [A or B]. Material from donor B then became unavailable, and patients were instead treated with stool from donor A or one of four other donors. At this point, donor B became available again, and '[t]he remaining participants allocated to active therapy all received FMT from donor B exclusively, as [the investigators] had not experienced any success with donor A'. In the original study, donor B's performance [seven of 18 treated patients achieved the primary end point] was compared against all other donors (two of 20 patients; odds ratio 5.5 [95% CI 1.0–44], mid-$p = 0.048$, Fisher's exact test).

In general, adaptive trial designs require special statistical methodologies.[48] To investigate whether typical statistical tests such as Fisher's exact test accurately measure the statistical significance of the data collected by Moayyedi et al., we simulated a simplified adaptive donor selection process. First, among the first 24 simulated patients, use all six donors four times each. Then select the best-performing donor to treat all the remaining 14 patients. Finally, compare the overall performance of this donor against all the others. In these simulations, if all the donors are actually the same [i.e. with equal probability $9/38 = 24\%$ of a positive patient outcome], the probability of finding mid-$p < 0.05$ is 8.8% [8843 of 10 000 simulations; 95% CI 8.7– 9.0%], nearly double the value that would be expected if the test were accurate for that type of data [i.e. the false positive rate, 5%]. The intuitive explanation is that, if all donors are the same, the selection process is merely setting aside donors who had 'bad luck' on their first patients, and they are not allowed to recover their performance with a run of 'good luck' later on. This is not a flaw of Fisher's exact test. Instead, it is problem of applying a statistical test to data that do not meet the assumptions of the test.

Although the adaptive procedure that led to extensive use of donor B may have improved the probability of the trial's success,[14] the procedure was not pre-specified, making it impossible to rigorously determine the degree to which the observed results are consistent with all donors being identical, that is, whether donor B simply had a 'lucky' initial run. Thus, the value mid-$p = 0.05$ reported above is inaccurately optimistic, and a *p*-value of 0.09 is a

Table 1. Summary of re-analyses

Data source	No. of patients	No. of donors	Evidence for donor superiority?[a]	Evidence for pool superiority?[a]	Association between donor α-diversity and outcomes?[b]	Separation in donor microbiota composition by patient outcome?[c]
Rossen et al.	23	15	No	—	—	—
Uygun et al.	30	10	No [0 p.p., 95% CI 0–69]	—	—	—
Vaughn et al.	15[d]	4	No [0 p.p., 95% CI 0–80]	—	—	—
Moayyedi et al.	38	6	No	—	—	—
Paramsothy et al.	78	14	No	No [0 p.p., 95% CI 0–46]	—	—
Costello et al.	38	19	No	No [0 p.p., 95% CI 0–74]	—	—
Jacob et al.	20	4	Yes [*p* = 0.02]	No [88 p.p., 95% CI 0–100]	No	No
Pool meta-analysis[e]	137	36	No	No [41 p.p., 95% CI 0–88]	—	—
Goyal et al.	21	21	—	—	No	No
Nusbaum et al.	7	7	—	—	No	No
Kump et al.	17	14	—	—	No	No
Vermeire et al.	14	14	—	—	Reported[f]	—
Nishida et al.	41	41	—	—	No	—

p.p.: Percentage points. CI: confidence interval.

[a]Tests of donor superiority and pool superiority were made using Fisher–Freeman–Halton or Fisher's exact test. Quantifications of pool effect are differences in clinical efficacy between a very high-quality pool [90th percentile] and a poor pool [10th percentile].

[b]Using Mann–Whitney *U* test or Kruskal–Wallis test.

[c]Using PERMANOVA on Bray–Curtis β-diversity.

[d]More patients were treated in the trial but patient–donor linkages were reported for only 15 patients.

[e]Includes Paramsothy et al., Costello et al. and Jacob et al..

[f]Underlying data were not available for re-analysis.

Table 2. Tests of donor superiority, by donor, for Rossen et al. Donor labels are arbitrary. 'Success' means the patient reached the primary end point; 'failure' means they did not

Donor	Patients treated using this donor		Patients treated not using this donor		Fisher's exact test	
	Success	Failure	Success	Failure	Odds ratio [95% CI]	Mid-*p* value
A	4	4	3	12	3.7 [0.55–29]	0.18
B	1	2	6	14	1.2 [0.03–18]	0.89
C	0	2	7	14	0.0 [0.0–8]	0.47

better representation of the likelihood of a donor having the kind of success observed in this trial.

3.3. Uygun et al. 2017

In this study,[29] 30 ulcerative colitis patients were treated with FMT. Ten donors provided stool: one donor provided stool for 12 patients, one donor for seven patients, one donor for four patients, and seven donors for one patient each. The outcome was clinical remission at 12 weeks. There was no evidence of a donor effect [*p* = 0.97, Fisher–Freeman–Halton test on a 2 × 10 table of patient outcomes by donor].

To develop an estimate of the strength of the donor effect, we used a logistic regression, modelling the donor as a random effect. This model estimated that the difference in efficacy between a very high-quality donor [90th percentile in terms of efficacy] and a poor donor [10th percentile] is exactly 0 percentage points [95% CI 0–69 percentage points]. In other words, the regression's best estimate of the variation in efficacy among pools is zero, even without applying

any penalizations for model complexity. Thus, similar to Rossen et al., the data do not provide evidence for a donor effect, but they are also not inconsistent with a strong effect, as the upper confidence limit on the strength of the donor effect is a 69 percentage point difference in efficacy rate between high-quality and poor donors.

3.4. Vaughn et al. 2016

In this study,[49] 15 Crohn's disease patients, treated with FMT from four donors, had sufficient reported data to analyse a donor effect. The outcome was clinical response at week 4. There was no evidence of a donor effect [*p* = 0.64, Fisher–Freeman–Halton test], and the estimated difference in efficacy between a very high-quality donor and a poor donor was 0 percentage points [95% CI 0–80 percentage points].

3.5. Paramsothy et al. 2017

In this study,[23] 41 ulcerative colitis patients were randomized to receive blinded FMT. Another 37 were randomized to placebo and later received open-label FMT. Each of the 78 patients received FMT

from one of 21 pools of donors. Each pool included material from three to seven donors, out of 14 total donors. The outcome in this analysis was clinical remission with endoscopic remission or response at 8 weeks after either the blinded or the open-label FMT.

The original publication compared the best performing individual donor against all the others (14 of 38 patients achieved the primary outcome, vs seven of 40 assigned to other donors; odds ratio 2.7 [95% CI 0.96–8.2], mid-p = 0.06, Fisher's exact test). This result, although near mid-p = 0.05, is not convincing when subjected to multiple hypothesis correction, as there are 14 relevant hypotheses to be tested, one per donor (Bonferroni-corrected p = 1.0, Benjamini–Hochberg false discovery rate [FDR] = 0.78). In terms of pools of donors, there was also no evidence that any particular pool was associated with better outcomes [p = 0.76, Fisher–Freeman–Halton test on a 2 × 21 table of patient outcomes by pool].

To develop an estimate of the strength of the donor effect, we used the same mixed logistic regression as for Uygun et al. but used pools rather than donors. As with Uygun et al., the estimated difference in efficacy between a very high-quality pool and a poor pool is exactly 0 percentage points [95% CI 0–46 percentage points].

A follow-up study[44] to that of Paramsothy et al. identified specific bacterial taxa associated with groups of pools that were linked to donors with positive or negative outcomes, but these results were not subject to multiple hypothesis correction.

3.6.   Costello et al. 2019

In this study,[45] 38 ulcerative colitis patients received FMT from one of 11 pools. Each pool included stool from three or four donors, out of 19 total donors. The outcome was steroid-free remission at 8 weeks. Similar to Paramsothy et al., there was no evidence of heterogeneity in patient outcomes by donor pool [p = 0.50, Fisher–Freeman–Halton test on 2 × 11 table] nor evidence of better outcomes for any particular donor [mid-p > 0.05 for all donors, Fisher's exact test]. Furthermore, a logistic regression estimated the difference between a very high-quality pool and a poor pool as 0 percentage points but with a wide confidence interval [95% CI 0–74 percentage points].

3.7.   Jacob et al. 2017

In this study,[24] 20 ulcerative colitis patients received FMT from one of six pools. Each pool included stool from two donors, out of four

total donors. In this analysis, the outcome was clinical response at 4 weeks. Curiously, although a test for differences in patient outcomes by pool was statistically significant [Fisher–Freeman–Halton test on 2 × 6 table, p = 0.02], a test for differences in outcomes by donor was not [mid-p > 0.05 for all four donors, Fisher's exact test]. A logistic regression estimated the difference in efficacy between very high-quality pools and low-quality pools as very large [88 percentage points] but not statistically significant [95% CI 0–100 percentage points].

There was no evidence for an association between patient outcomes and the bacterial community α-diversity of the pool they received [Figure 1a; p = 0.39, Mann–Whitney U test]. In fact, more diverse pools were associated with worse patient outcomes. In our analysis, donor microbiota compositions also did not separate according to their associated patient outcomes [Figure 1b; p = 0.077, PERMANOVA], while they did in the original analysis [figure 4C in that publication, p = 0.044], probably due to the differences in the 16S rRNA gene sequencing data processing pipeline. Our analysis used de-noising and the Bray–Curtis β-diversity metric, while Jacob et al. used 97% operational taxonomic units and the UniFrac metric.

3.8.   Meta-analysis of pool studies

Paramsothy et al., Costello et al. and Jacob et al. have sufficiently similar designs and the available data to allow a meta-analysis. A logistic regression on the combined data from all three studies, with study as a fixed effect, estimated the difference in efficacy between very high-quality pools and low-quality pools as 41 percentage points but not statistically significant [95% CI 0–88 percentage points].

3.9.   Goyal et al. 2018

In this study,[28] 21 patients with Crohn's disease, ulcerative colitis or indeterminate colitis received FMT and were available for follow-up. We used remission at 30 days as the outcome. Each patient received stool from a different donor, precluding any test of donor superiority. However, 16S rRNA gene sequencing was performed, allowing for an analysis of donation characteristic superiority, similar to Jacob et al. above. Similar to that study, there was no evidence for an association between patient outcomes and the α-diversity of donor stool
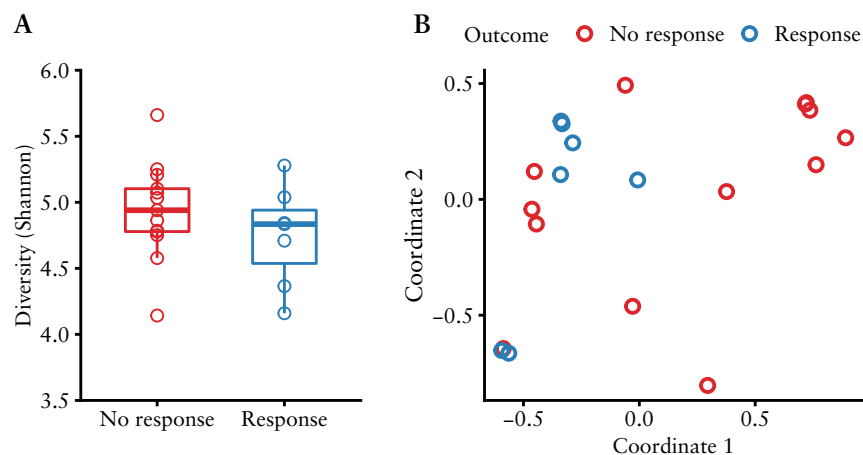
Figure 1. Microbiome analyses for Jacob et al. [a] Bacterial α-diversity of FMT pools does not differ significantly by patient outcomes [p = 0.39, Mann–Whitney U test]. Each point represents a single patient. [b] Donor microbiota compositions [points; x- and y-coordinates show multidimensional scaling using Bray–Curtis dissimilarity] do not statistically significantly separate by associated patient outcome [colour; p = 0.076, PERMANOVA]. Each point represents the composition of the pool used to treat an individual patient.

[Figure 2a; $p = 0.97$, Mann–Whitney U test] nor for separation in donor microbiota compositions [Figure 2b; $p = 0.7$, PERMANOVA].

3.10. Nusbaum et al. 2018

In this study,[36] seven ulcerative colitis patients were treated with FMT, each using a different donor. There was no evidence for an association between patient outcomes and α-diversity of donor stool [$p = 0.86$, Mann–Whitney U test] nor for separation in donor microbiota compositions [$p = 0.36$, PERMANOVA].

3.11. Kump et al. 2017

In this study,[27] 17 ulcerative colitis patients received five FMTs from one of 14 donors. Twelve donors were used in one patient each, one donor was used in two patients and one donor was used in three patients. As in Goyal et al., this distribution of patients across donors does not permit an assessment of donor superiority. However, like Jacob et al. and Goyal et al., this study tested whether increased α-diversity was associated with improved patient outcomes.

We detected no difference in diversity by patient outcome [Figure 3a; $p = 0.39$, Kruskal–Wallis test; $p > 0.05$ for all three comparisons, Mann–Whitney U test], nor did we detect separation in donor microbiota according to associated patient outcomes [Figure 3b; $p = 0.15$, PERMANOVA].

The original publication reported statistically significant differences between the α-diversity of the 16 available donor stool samples associated with the four patients who achieved remission compared to the diversity of 12 available samples associated with the four patients who did not respond to FMT. Our analysis has a different conclusion, finding no support for a donor effect, principally because we analysed the data as only eight independent data points, that is, one data point for each patient [four remission, four no response]. By contrast, Kump et al. analysed the data as if there were 28 independent data points, that is, one data point for each donation [12 associated with remission, 16 associated with no response]. Using donations rather than patients to determine N is a statistical error known as 'pseudoreplication'.[50] Repeated measurements of the material delivered to a patient do not constitute independent measures of the patient's outcome. In other words, when

testing for an association between patient outcomes and a donation characteristic such as α-diversity, the weight of the evidence is determined by the number of patient outcomes, not by the number of microbiome measurements.[51]

3.12. Vermeire et al. 2016

In this study,[33] 14 patients with Crohn's disease or ulcerative colitis received FMT, each from a different donor. The original study reported that the bacterial α-diversity of donations varied by patient outcomes, namely that patients with successful outcomes received more bacterially diverse donations [Mann–Whitney U test, $p = 0.012$]. However, the underlying bacterial sequencing data were not available for re-analysis.

3.13. Nishida et al. 2017

In this study,[26] 41 patients with ulcerative colitis received FMT, each from a different donor. The outcome was clinical response at 8 weeks. The original study found no difference in the α-diversity of the stool from seven donors whose patients met the primary end point, compared to 19 donors whose patients did not [$p = 0.69$]. The study also tested whether the abundance of ten taxa were differentially abundant among those two groups of donors, finding p-values less than 0.05 for two of nine taxa, neither of which is statistically significant after Benjamini–Hochberg multiple hypothesis correction. The underlying bacterial sequencing data were not available for re-analysis.

4. Discussion

In this study, we re-evaluated 12 studies that used FMT for IBD. In ten of 12 cases, there was no statistically significant evidence of a donor effect when using our statistical approach and bioinformatic pipeline. In one of the remaining studies, Jacob et al., there was a statistically significant difference in efficacy by pool of donor stool, but not by donor, nor by the α-diversity of the stool or by donor microbiota composition. We were unable to re-analyse the last study, Vermeire et al., which reported a statistically significant effect.

The major strengths of this study were the rigorous statistical approach used, which avoided pseudoreplication and misinterpretation
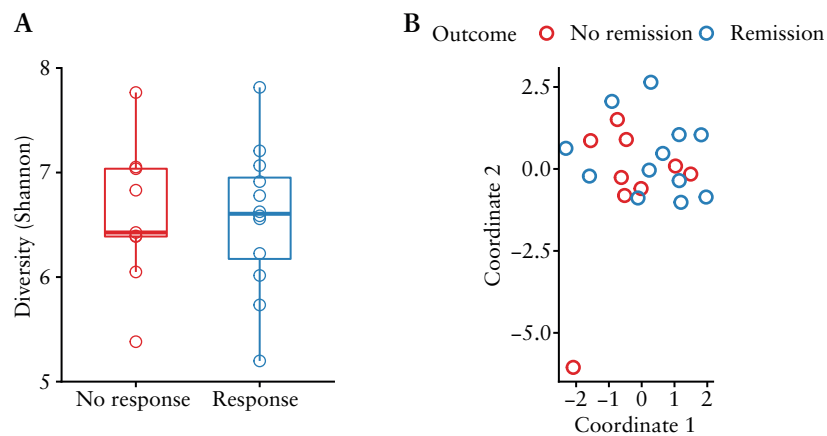
Figure 2. Microbiome analyses for Goyal et al. [a] Bacterial α-diversity of FMT pools does not differ significantly by patient outcomes [$p = 0.97$, Mann–Whitney U test]. Each point represents a single patient. [b] Donor microbiota compositions [points; x- and y-coordinates show multidimensional scaling using Bray–Curtis dissimilarity] do not statistically significantly separate by associated patient outcome [colour; $p = 0.70$, PERMANOVA]. Each point represents the composition of the first sample from the donor used to treat an individual patient.
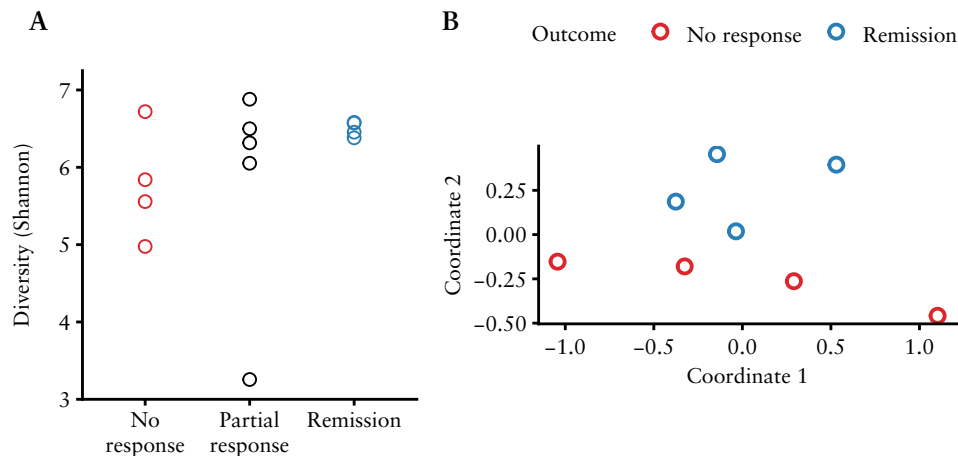
Figure 3. Microbiome analyses for Kump et al. [a] Bacterial α-diversity does not differ by patient outcome [$p = 0.39$, Kruskal–Wallis test; $p > 0.05$ for all three comparisons, Mann–Whitney U test]. Each point represents a single patient. The diversity shown for each patient [vertical axis] is the average of the diversity of the sampled FMTs used in that patient. [b] Donor microbiota does not statistically significantly separate by associated patient outcome [colour; $p = 0.15$, PERMANOVA]. Each point represents the composition of the first donor sample administered to each patient.

of results from an adaptive trial design, and the uniformity of the re-analysis of the 16S rRNA microbiota composition data.

However, this study is also subject to four key limitations. First, our analysis considered only donor effects and made no reference to patient factors such as disease history. Given the small sample sizes, adding patient factors as predictors would only decrease our ability to detect donor effects, which was contrary to our main aim. We do not mean to imply that all IBD patients are the same nor should that they receive the same kind of treatment. In fact, a principal conclusion from this study is that patient factors, not donor factors, are the main determinants of an FMT patient's clinical outcome.

Second, our analysis considered only a subset of possible donor effects. We performed re-analyses only for donor/pool superiority [e.g. donor A vs donor B] and material characteristic superiority, two of the four types of donor effect we enumerated. We did not re-analyse the third type of donor effect, donor characteristic superiority, because most studies report only little information about the donors.[52] Only one re-analysed study had tested for donor characteristic superiority: Goyal et al. found no association between donor age and patient outcome. A re-analysis of the final type of donor effect, patient–donor matching, was beyond the scope of this study, and no study reported a rigorous analysis of such a matching. It remains possible that this type of donor effect will be uncovered at some point in the future, but we caution that donor–patient matching presents important challenges, as the number of possible donor and patient factors that could be matched, and the number of ways the levels of those factors corresponds to a favourable or unfavourable match, are very large.

Third, we considered only a single type of donation characteristic superiority, focusing on 16S rRNA microbiota composition α-diversity and microbiota composition β-diversity. More types of analysis could be performed using 16S rRNA data, but previous work suggests that identifying individual taxa associated with effective donors would require large trials.[15] Thus, patient biomarkers, such as mucosal immune markers, may be more likely to be able to detect donor effects.[30] We caution, however, that attempts to find associations between every conceivable donor factor, patient factor and patient outcome risk 'p-hacking'.[53]

As a final limitation, we note that, although we aimed to include all IBD studies, most studies focused on ulcerative colitis.

Only one study considered Crohn's patients alone, and two studies included both ulcerative colitis and Crohn's disease patients. Only one study of ulcerative colitis and Crohn's disease may be subject to donor effects, or they may be subject to different kinds of effects, and there was insufficient available data to disentangle these possibilities.

Overall, we found only very weak evidence for donor effects. On the other hand, the size of these studies is small relative to the number of patients required to reliably detect a plausible donor effect.[15,30] The available data therefore cannot definitively rule out a large donor effect, and it remains undetermined if differences between donors are clinically relevant to FMT for IBD. We propose that larger studies, with careful biomarker selection and transparent reporting of results,[36] are the best direction forward to characterizing the donor effect for FMT in IBD.

Conflict of Interest

Acknowledgments

Author Contributions

S.W.O. is the guarantor. S.W.O. conceived the study, obtained data, performed the analysis, and wrote the manuscript. Y.G. obtained data; improved the analysis, interpretation, and presentation of the results; and revised the manuscript. All authors approved the final version of the manuscript.

Funding

Supplementary Data

Supplementary data are available at *ECCO-JCC* online.

References

1. Marchesi JR, Adams DH, Fava F, *et al*. The gut microbiota and host health: a new clinical frontier. *Gut* 2016;**65**:330–9.
2. Allegretti JR, Mullish BH, Kelly C, Fischer M. The evolution of the use of faecal microbiota transplantation and emerging therapeutic indications. *Lancet* 2019;**394**:420–31.
3. McDonald LC, Gerding DN, Johnson S, *et al*. Clinical Practice Guidelines for *Clostridium difficile* Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clin Infect Dis* 2018;**66**:e1–e48.
4. Debast SB, Bauer MP, Kuijper EJ; European Society of Clinical Microbiology and Infectious Diseases. European Society of Clinical Microbiology and Infectious Diseases: update of the treatment guidance document for *Clostridium difficile* infection. *Clin Microbiol Infect* 2014;**20** Suppl 2:1–26.
5. McDonald JAK, Mullish BH, Pechlivanis A, *et al*. Inhibiting growth of *Clostridioides difficile* by restoring valerate, produced by the intestinal microbiota. *Gastroenterology* 2018;**155**:1495–1507.e15.
6. Mullish BH, McDonald JAK, Pechlivanis A, *et al*. Microbial bile salt hydrolases mediate the efficacy of faecal microbiota transplant in the treatment of recurrent *Clostridioides difficile* infection. *Gut* 2019;**68**:1791–800.
7. Olesen SW, Panchal P, Chen J, Budree S, Osman M. Global disparities in faecal microbiota transplantation research. *Lancet Gastroenterol Hepatol* 2020;**5**:241.
8. Lopetuso LR, Ianiro G, Allegretti JR, *et al*. Fecal transplantation for ulcerative colitis: current evidence and future applications. *Expert Opin Biol Ther* 2020;**20**:343–51.
9. David LA, Maurice CF, Carmody RN, *et al*. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 2014;**505**:559–63.
10. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature* 2007;**449**:804–10.
11. Rossen NG, Fuentes S, van der Spek MJ, *et al*. Findings from a randomized controlled trial of fecal transplantation for patients with ulcerative colitis. *Gastroenterology* 2015;**149**:110–118.e4.
12. Wilson BC, Vatanen T, Cutfield WS, O'Sullivan JM. The super-donor phenomenon in fecal microbiota transplantation. *Front Cell Infect Microbiol* 2019;**9**:2.
13. Olesen SW, Leier MM, Alm EJ, Kahn SA. Searching for superstool: maximizing the therapeutic potential of FMT. *Nat Rev Gastroenterol Hepatol* 2018;**15**:387–8.
14. Olesen SW, Gurry T, Alm EJ. Designing fecal microbiota transplant trials that account for differences in donor stool efficacy. *Stat Methods Med Res* 2018;**27**:2906–17.
15. Duvallet C, Zellmer C, Panchal P, Budree S, Osman M, Alm EJ. Framework for rational donor selection in fecal microbiota transplant clinical trials. *PLoS One* 2019;**14**:e0222881.
16. Barnes D, Ng K, Smits S, Sonnenburg J, Kassam Z, Park KT. Competitively selected donor fecal microbiota transplantation: butyrate concentration and diversity as measures of donor quality. *J Pediatr Gastroenterol Nutr* 2018;**67**:185–7.
17. Fuentes S, Rossen NG, van der Spek MJ, *et al*. Microbial shifts and signatures of long-term remission in ulcerative colitis after faecal microbiota transplantation. *ISME J* 2017;**11**:1877–89.
18. Ng SC, Kamm MA, Yeoh YK, *et al*. Scientific frontiers in faecal microbiota transplantation: joint document of Asia-Pacific Association of Gastroenterology (APAGE) and Asia-Pacific Society for Digestive Endoscopy (APSDE). *Gut* 2020;**69**:83–91.
19. Osman M, Abend A, Panchal P, Kassam Z, Budree S. Does the donor matter? Microbiome sequencing to evaluate lower donor efficacy in fecal microbiota transplantation for recurrent *Clostridium difficile* infection. *Gastroenterology* 2018;**154**:S-25–S-26.
20. Ray A, Jones C. Does the donor matter? Donor vs patient effects in the outcome of a next-generation microbiota-based drug trial for recurrent *Clostridium difficile* infection. *Future Microbiol* 2016;**11**:611–6.
21. Zuo T, Wong SH, Cheung CP, *et al*. Gut fungal dysbiosis correlates with reduced efficacy of fecal microbiota transplantation in *Clostridium difficile* infection. *Nat Commun* 2018;**9**:3663.
22. Moayyedi P, Surette MG, Kim PT, *et al*. Fecal microbiota transplantation induces remission in patients with active ulcerative colitis in a randomized controlled trial. *Gastroenterology* 2015;**149**:102–109.e6.
23. Paramsothy S, Kamm MA, Kaakoush NO, *et al*. Multidonor intensive faecal microbiota transplantation for active ulcerative colitis: a randomised placebo-controlled trial. *Lancet* 2017;**389**:1218–28.
24. Jacob V, Crawford C, Cohen-Mekelburg S, *et al*. Single delivery of high-diversity fecal microbiota preparation by colonoscopy is safe and effective in increasing microbial diversity in active ulcerative colitis. *Inflamm Bowel Dis* 2017;**23**:903–11.
25. Paramsothy S, Paramsothy R, Rubin DT, *et al*. Faecal microbiota transplantation for inflammatory bowel disease: A systematic review and meta-analysis. *J Crohns Colitis* 2017;**11**:1180–99.
26. Nishida A, Imaeda H, Ohno M, *et al*. Efficacy and safety of single fecal microbiota transplantation for Japanese patients with mild to moderately active ulcerative colitis. *J Gastroenterol* 2017;**52**:476–82.
27. Kump P, Wurm P, Gröchenig HP, *et al*. The taxonomic composition of the donor intestinal microbiota is a major factor influencing the efficacy of faecal microbiota transplantation in therapy refractory ulcerative colitis. *Aliment Pharmacol Ther* 2018;**47**:67–77.
28. Goyal A, Yeh A, Bush BR, *et al*. Safety, clinical response, and microbiome findings following fecal microbiota transplant in children with inflammatory bowel disease. *Inflamm Bowel Dis* 2018;**24**:410–21.
29. Uygun A, Ozturk K, Demirci H, *et al*. Fecal microbiota transplantation is a rescue treatment modality for refractory ulcerative colitis. *Medicine (Baltimore)* 2017;**96**:e6479.
30. Olesen SW. Power calculations for detecting differences in efficacy of fecal microbiota donors. *medRxiv* 2020. doi:10.1101/2020.04.16.20068361.
31. Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond 'p < 0.05'. *Am. Stat*. 2019;**73**:1–19.
32. Budree S, Tu E, Leith T, *et al*. The association of stool donor diet on microbial profile and clinical outcomes of fecal microbiota transplantation in *Clostridium difficile* infection. *Gastroenterology* 2017;**152**:S630–S631.
33. Vermeire S, Joossens M, Verbeke K, *et al*. Donor species richness determines faecal microbiota transplantation success in inflammatory bowel disease. *J Crohns Colitis* 2016;**10**:387–94.
34. Kazerouni A, Wein LM. Exploring the efficacy of pooled stools in fecal microbiota transplantation for microbiota-associated chronic diseases. *PLoS One* 2017;**12**:e0163956.
35. Pigneur B, Sokol H. Fecal microbiota transplantation in inflammatory bowel disease: the quest for the holy grail. *Mucosal Immunol* 2016;**9**:1360–5.
36. Nusbaum DJ, Sun F, Ren J, *et al*. Gut microbial and metabolomic profiles after fecal microbiota transplantation in pediatric ulcerative colitis patients. *FEMS Microbiol. Ecol*. 2018;**94**:fiy133.
37. Khanna S, Vazquez-Baeza Y, González A, *et al*. Changes in microbial ecology after fecal microbiota transplantation for recurrent *C. difficile* infection affected by underlying inflammatory bowel disease. *Microbiome* 2017;**5**:55.
38. R Core Team. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, 2019.
39. Fay MP. Confidence intervals that match Fisher's exact or Blaker's exact tests. *Biostatistics* 2010;**11**:373–4.
40. Mehta CR, and Hilton JF. Exact power of conditional and unconditional tests: Going beyond the 2x2 contingency table. *Am. Stat*. 1993. doi:10.1080/00031305.1993.10475946.
41. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 2001;**26**:32–46.
42. Oksanen J, Blanchet FG, Friendly M, *et al*. *vegan: Community Ecology Package*. https://cran.r-project.org/package=vegan. Accessed 1 August 2020. 2019.

43. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *J. Stat. Softw*. 2015. doi:10.18637/jss.v067.i01.

44. Paramsothy S, Nielsen S, Kamm MA, *et al*. Specific bacteria and metabolites associated with response to fecal microbiota transplantation in patients with ulcerative colitis. *Gastroenterology* 2019;**156**:1440–1454.e2.

45. Costello SP, Conlon MA, Andrews JM. Fecal microbiota transplantation for ulcerative colitis – reply. *JAMA* 2019;**321**:2240–1.

46. Bolyen E, Rideout JR, Dillon MR, *et al*. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;**37**:852–7.

47. Amir A, McDonald D, Navas-Molina JA, *et al*. Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* 2017;**2**:e00191–16.

48. Bhatt DL, Mehta C. Adaptive designs for clinical trials. *N Engl J Med* 2016;**375**:65–74.

49. Vaughn BP, Vatanen T, Allegretti JR, *et al*. Increased intestinal microbial diversity following fecal microbiota transplant for active Crohn's disease. *Inflamm Bowel Dis* 2016;**22**:2182–90.

50. Hurlbert SH. Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr* 1984;**54**:187–211.

51. Lazic SE, Clarke-Williams CJ, Munafò MR. What exactly is 'N' in cell culture and animal experiments? *PLoS Biol* 2018;**16**:e2005282.

52. Bafeta A, Yavchitz A, Riveros C, Batista R, Ravaud P. Methods and reporting studies assessing fecal microbiota transplantation: a systematic review. *Ann Intern Med* 2017;**167**:34–9.

53. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. *PLoS Biol* 2015;**13**:e1002106.