Title: Re-evaluating the evidence for fecal microbiota transplantation “super-donors” in inflammatory bowel disease

Short title: Re-evaluating FMT “super-donors”

Authors: Scott W. Olesen, PhD[1,2]; Ylaine Gerardin, PhD[3]

1 OpenBiome, Cambridge, MA
2 Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA
3 Finch Therapeutics, Somerville, MA

Correspondence
Scott Olesen
2067 Massachusetts Avenue
Cambridge, MA 02140
solesen@openbiome.org

Keywords: fecal microbiota transplantation, inflammatory bowel disease, donor effect, super donor

Main text word count: 3880

SUMMARY (200 words)

Background: Fecal microbiota transplantation (FMT) is a recommended treatment for recurrent *Clostridioides difficile* infection, and there is promise that FMT may be effective for conditions like 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.

Aim & Methods: Here we re-evaluate evidence for donor effects in published FMT clinical trials for IBD.

Results: In 7 of 9 published studies, no statistically significant donor effect was detected when rigorously re-evaluating the original analyses. One study had 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 effect and very substantial donor effects.

Conclusion: Although we found very little evidence for donor effects, the existing data also cannot rule out the possibility that donor effects are clinically important. Large clinical trials prospectively designed to detect donor effects are likely necessary to determine if donor effects are clinically relevant for IBD.

INTRODUCTION

The human microbiome is increasingly understood to play a key role in health and disease.[1] Fecal 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 are still developing, FMT is being investigated as a therapy for dozens of other microbiome-related indications.[2,5,6]

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.[7,8] 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 like "donor effect", "super-donor", and "super-stool".[6,9–12]

If FMT's efficacy varied widely across stools or donors, then rational selection of FMT material based on biomarkers predictive of efficacy could improve the clinical practice of FMT.[6,11,13,14] Differences in FMT efficacy between stool donors or specific stools could be also an important starting point for scientific investigations into the "active ingredient"

in FMT donations.[11] It is therefore important to quantify the extent of donor variability in indications where FMT is a promising treatment.

Although there is little evidence of a donor effect in the context of *C. difficile* infection,[15–17] the condition for which the use of FMT is best studied, multiple studies using FMT to treat inflammatory disease have tested for donor effects,[2,9,10,18–24] and some have reported statistically significant results. However, there are reasons to be skeptical of the clinical implications of these intriguing findings.

First, a previous report[25] 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 it 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 analyzed donor effects by testing null hypotheses that are biologically unlikely: either that all donors produce stool with exactly equal clinical

69 efficacy, or that particular features of stool, like its bacterial community diversity, are
70 completely unrelated to clinical efficacy. However, it is highly improbable that all stool
71 has precisely the same therapeutic effect: if a study is large enough, the difference
72 between two donors' treatment efficacies would almost certainly become statistically
73 significant as the number of treated patients increases. We therefore propose that,
74 rather than merely asking if donor effects exist or not, we should ask about the
75 magnitude of the donor effect and whether it is clinically relevant. In other words, rather
76 than merely testing for the statistical significance of donor effects, we should also
77 assess effect sizes.[26]
78
79 Here we re-evaluate the existing evidence for donor effects and discuss the implications
80 of that evidence for clinical trial design and clinical practice. First, we establish an
81 ontology of the various concepts referred to as "donor effects". Second, we lay out a
82 rigorous framework for identifying and quantifying donor effects. Third, we re-evaluate
83 the existing literature using the conceptual ontology and the rigorous framework. Finally,
84 we discuss the implications of this re-evaluation for future FMT research in IBD.
85
86
87 **METHODS**
88
89 **Distinguishing types of stool superiority**
90 Discussions of "super-donors" and "super-stool" have suggested that stool might be
91 superior in at least 4 distinct but conceptually related senses. To avoid confusion in our

92 re-evaluation of the evidence for donor effects, we distinguish between these definitions
93 of stool superiority:
94
95 1. *Donor superiority*, in which particular donors are associated with better clinical
96 outcomes for the recipient patients. For example, Moayyedi *et al.*[18] tested
97 whether a particular donor ("donor B") was associated with better patient
98 outcomes, compared to the other donors.
99 2. *Donor characteristic superiority*, in which donors with particular characteristics
100 are associated with better outcomes. For example, donor age, diet, and host
101 genetics have been suggested as potential factors in donor superiority.[10,24,27]
102 3. *Material characteristic superiority*, in which stool donations, pools of stool, or
103 other fecal microbiota preparations that have some particular characteristic are
104 associated with better outcomes. For example, Vermeire *et al.*[28] tested whether
105 stools with higher bacterial diversity are associated with better outcomes.
106 4. *Donor-recipient match superiority*, in which certain combinations of donors and
107 recipients are associated with better outcomes,[9,10,21,29] analogous to how donors'
108 and recipients' blood types are matched for blood transfusions. For example,
109 FMT studies have tested whether stool from "related" donors,[30] typically defined
110 as first-degree relatives but also sometimes including spouses or partners, is
111 associated with better or worse patient outcomes.
112
113 These types of superiority are conceptually related and not mutually exclusive. For
114 example, imagine it were the case that female donors were associated with better

115 patient outcomes than male donors, consistent with definition 2, “donor characteristic
116 superiority”. However, because donors have a clinical effect on the patient only via their
117 donation, sex of donor cannot be the molecular mechanism by which some donations
118 are more efficacious than others. It would have to be that sex determined or correlated
119 with some component of the stool that made those donations more effective. In other
120 words, there must be an underlying “material characteristic superiority”, definition 3, that
121 correlates with donor sex. By a similar argument, even if donor-recipient match
122 superiority is the most accurate model, simpler types of superiority may be more
123 parsimonious.[12]
124
125 **Identifying and quantifying donor effects**
126 We used a two-step framework for characterizing donor effects. First, we used statistical
127 tests to characterize the strength of the evidence for the reported type of donor effect.
128 Statistics were computed using the R programming language[31] (version 3.6.0). To test
129 the hypothesis that there is any difference in efficacy between donors (or pools of
130 donors), we used the Fisher-Freeman-Halton test (function *fisher.test*) on 2 × *D*
131 contingency tables, where *D* is the number of donors (or pools). For 2 × 2 contingency
132 tables (e.g., to test if one donor has a different treatment efficacy than all the others), we
133 used Fisher’s exact test (function *exact2x2*)[32] and the mid-*p* value to prevent the tests
134 from being overly conservative.[33] To test for material (i.e., donation or pool)
135 characteristic superiority, we used the Mann-Whitney *U* test (*wilcox.test*) or Kruskal-
136 Wallis test (*kruskal.test*). When 16S rRNA gut microbiota composition data were

137 available (see below), we tested for separation in donor microbiota compositions using
138 PERMANOVA[34] (function *Adonis* in the R *vegan* package).[35]
139
140 Second, whether or not an effect was detected, we use a random effect logistic
141 regression (function *glmer* in the R *lme4* package)[36] to quantify the effect size. The
142 regression model assumes that the log odds of the efficacy of different donors (or pools
143 of stool) is Gaussian distributed. The regression returns the standard deviation of that
144 distribution, which is a measure of the "spread" in donors' efficacies. For clarity, we
145 convert the regression's estimates into the absolute difference in efficacy rates between
146 a 90th percentile donor (a very high quality donor) and a 10th percentile donor (a poor
147 donor). For example, an effect size of 30 percentage points means that, if a low quality
148 donor is 10% efficacious, then a high quality donor would be 40% efficacious.
149
150 **Studies re-analyzed**
151 In the re-analysis, we included interventional FMT clinical studies that tested the
152 hypothesis that one donor or donation characteristic was associated with improved IBD
153 patient outcomes. Studies were identified using citations from an existing systematic
154 review,[21] narrative review,[10] and expert opinion.[6] For each study, we considered only a
155 single dichotomous outcome, whether a patient achieved the trial's primary clinical
156 endpoint. When 16S rRNA gut microbiota composition data were available, we tested
157 for associations between donor material α-diversity and patient outcomes,[23,24,28] since
158 part of the motivation for using pools of stool from multiple donors is the hypothesis that
159 high bacterial diversity in FMT material is beneficial.[13,19,29,37] When possible, we also

160 searched for separation in the microbiota compositions of donors according to the
161 outcomes of their associated patients, following the example of Jacob *et al*.[20]
162
163 **Microbiome data preparation**
164 We re-analyzed raw 16S rRNA sequencing data from three studies (Kump *et al.*, study
165 accession PRJEB11841; Jacob *et al.*, accession PRJNA388210; Goyal *et al.*, accession
166 PRJNA380944) using QIIME 2[38] (version 2019.7) and Deblur.[39] For Kump *et al.* and
167 Jacob *et al.*, paired reads were joined (*vsearch* plugin, default parameters), quality
168 filtered (*quality-filter* plugin, default parameters), and denoised using Deblur (trim length
169 253 nucleotides, 1 minimum reads, otherwise default parameters). For Goyal *et al.*, the
170 sequencing data was single-ended (no joining required) and the trim length was 150
171 nucleotides. Alpha-diversity was computed by down-sampling all samples to the
172 minimum number of denoised read counts across samples and then using the Shannon
173 metric (*diversity* plugin). When a donor had multiple associated samples, we associated
174 a single α-diversity with each donor by computing the mean α-diversity over that donor's
175 samples. Beta-diversity, used for the PERMANOVA tests, was computed with the Bray-
176 Curtis metric. When a donor had multiple samples, we used only the first sample from
177 each donor.
178
179 **Code and data availability**
180 Computer code and underlying data to reproduce the results are online (DOI:
181 10.5281/zenodo.3780184, https://www.github.com/openbiome/donor-effects).
182

183

184 **RESULTS**

185

186 The results of the re-evaluation of the evidence for a donor effect in IBD from a

187 selection of FMT clinical studies in summarized in Table 1.

188

189 **Rossen *et al*. 2015**

190 In this study,[9] 23 ulcerative colitis patients were treated with FMT. 17 patients received

191 stool from 1 of 15 donors; 6 patients received stool from 2 different donors. The

192 outcome was clinical remission and endoscopic response at 12 weeks. In the original

193 publication, the investigators reported on the performance of 3 donors (Table 2). There

194 was no evidence of a donor effect (mid-$p > 0.05$, Fisher's exact test) for any of the 3

195 donors. However, the study is small enough that the data are not inconsistent with large

196 variability between donors. For example, the upper 95% confidence interval on donor

197 A's odds ratio of successful patient outcome compared to all other donors is 29. In other

198 words, the data do not provide evidence of a donor effect, but they also cannot

199 definitively rule out a strong donor effect.

200

201 **Moayyedi *et al*. 2015**

202 In this study,[18] 38 ulcerative colitis patients were treated with FMT. Each patient was

203 treated with stool from 1 of 6 donors. The outcome was remission at 7 weeks. Donors

204 were used according to an adaptive process, described in the original publication.

205 Briefly, at the start of the trial, patients were treated with stool from one of two donors (A

206 or B). Material from donor B then became unavailable, and patients were instead
207 treated with stool from donor A or 1 of 4 other donors. At this point, donor B became
208 available again, and "[t]he remaining participants allocated to active therapy all received
209 FMT from donor B exclusively, as [the investigators] had not experienced any success
210 with donor A". In the original study, donor B's performance (7 of 18 treated patients
211 achieved the primary endpoint) was compared against all other donors (2 of 20 patients;
212 odds ratio 5.5 [95% CI 1.0 to 44], mid-p = 0.048, Fisher's exact test).
213
214 In general, adaptive trial designs require special statistical methodologies.[40] To
215 investigate whether typical statistical tests like Fisher's exact test accurately measure
216 the statistical significance of the data collected in Moayyedi *et al*., we simulated a
217 simplified adaptive donor selection process. First, among the first 24 simulate patients,
218 use all 6 donors 4 times each. Then select the best-performing donor to treat all the
219 remaining 14 patients. Finally, compare the overall performance of this donor against all
220 the others. In these simulations, if all the donors are actually the same (i.e., with equal
221 probability 9/38 = 24% of a positive patient outcome), the probability of finding mid-p <
222 0.05 is 8.8% (8843 of 10,000 simulations; 95% CI 8.7% to 9.0%), nearly double the
223 value that would be expected if the test were accurate for that type of data (i.e., the
224 false positive rate, 5%). The intuitive explanation is that, if all donors are the same, the
225 selection process is merely setting aside donors who had "bad luck" on their first
226 patients, and they are not allowed to recover their performance with a run of "good luck"
227 later on. This is not a flaw of Fisher's exact test. Instead, it is problem of applying a
228 statistical test to data that do not meet the assumptions of the test.

229

230 Although the adaptive procedure that led to extensive use of donor B may have
231 improved the probability of the trial's success,[12] the procedure was not pre-specified,
232 making it impossible to rigorously determine the degree to which the observed results
233 are consistent with all donors being identical, that is, whether donor B simply had a
234 "lucky" initial run. Thus, the value mid-p = 0.05 reported above is inaccurately optimistic,
235 and a p-value of 0.09 is a better representation of the likelihood of a donor having the
236 kind of success observed in this trial.

237

238 **Paramsothy *et al*. 2017**

239 In this study,[19] 41 ulcerative colitis patients were randomized to receive blinded FMT.
240 Another 37 were randomized to placebo and later received open-label FMT. Each of the
241 78 patients received FMT from 1 of 21 pools of donors. Each pool included material
242 from 3 to 7 donors, out of 14 total donors. The outcome in this analysis was clinical
243 remission with endoscopic remission or response at 8 weeks after either the blinded or
244 open-label FMT.

245

246 The original publication compared the best performing individual donor against all the
247 others (14 of 38 patients achieved the primary outcome, versus 7 of 40 assigned to
248 other donors; odds ratio 2.7 [95% CI 0.96 to 8.2], mid-p = 0.06, Fisher's exact test). This
249 result, although near mid-p = 0.05, is not convincing when subjected to multiple
250 hypothesis correction, as there are 14 relevant hypotheses to be tested, one per donor
251 (Bonferroni-corrected p = 1.0, Benjamini-Hochberg FDR = 0.78). In terms of pools of

252 donors, there was also no evidence that any particular pool was associated with better
253 outcomes (p = 0.76, Fisher-Freeman-Halton test on 2 × 21 table of patient outcomes by
254 pool).
255
256 To develop an estimate of the strength of the donor effect, we used a logistic
257 regression, modeling the pool as a random effect. This model estimated that the
258 difference in efficacy between a very high quality pool (90th percentile in terms of
259 efficacy) and a poor pool (10th percentile) is exactly 0 percentage points (95% CI 0 to
260 46 percentage points). In other words, the regression's best estimate of the variation in
261 efficacy among pools is zero, even without applying any penalizations for model
262 complexity. Thus, similar to Rossen *et al.*, the data do not provide evidence for a donor
263 effect, but they are also not inconsistent with a strong effect, since the upper confidence
264 limit on the strength of the donor effect is a 46 percentage point difference in efficacy
265 rate between high- and low-quality donors.
266
267 **Costello *et al*. 2019**
268 In this study,[37] 38 ulcerative colitis patients received FMT from 1 of 11 pools. Each pool
269 included stool from 3 or 4 donors, out of 19 total donors. The outcome was steroid-free
270 remission at 8 weeks. Similar to Paramsothy *et al*., there was no evidence of
271 heterogeneity in patient outcomes by donor pool (p = 0.50, Fisher-Freeman-Halton test
272 on 2 × 11 table) nor evidence of better outcomes for any particular donor (mid-p > 0.05
273 for all donors, Fisher's exact test). Furthermore, a logistic regression estimated the
274 difference between a very high quality pool (90th percentile in terms of efficacy) and low

275 quality pool (10th percentile) as 0 percentage points but with a wide confidence interval
276 (95% CI 0 to 74 percentage points).
277
278 **Jacob *et al*. 2017**
279 In this study,[20] 20 ulcerative colitis patients received FMT from 1 of 6 pools. Each pool
280 included stool from 2 donors, out of 4 total donors. In this analysis, the outcome was
281 clinical response at 4 weeks. Curiously, although a test for differences in patient
282 outcomes by pool was statistically significant (Fisher-Freeman-Halton test on 2 × 6
283 table, $p = 0.02$), a test for differences in outcomes by donor was not (mid-$p > 0.05$ for all
284 4 donors, Fisher's exact test). A logistic regression estimated the difference in efficacy
285 between very high quality pools and low quality pools as very large (88 percentage
286 points) but not statistically significant (95% confidence interval 0 to 100 percentage
287 points).
288
289 There was no evidence for an association between patient outcomes and the bacterial
290 community α-diversity of the pool they received (Figure 1a; $p = 0.39$, Mann-Whitney U
291 test). In fact, more diverse pools were associated with worse patient outcomes. In our
292 analysis, donor microbiota compositions also did not separate according to their
293 associated patient outcomes (Figure 1b; $p = 0.077$, PERMANOVA), while they did in the
294 original analysis (Figure 4C in that publication, $p = 0.044$), likely due to the differences in
295 the 16S rRNA sequencing data processing pipeline. Our analysis used de-noising and
296 the Bray-Curtis β-diversity metric, while Jacob *et al.* used 97% operational taxonomic
297 units and the UniFrac metric.

298

299 **Meta-analysis of pool studies**

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

305

306 **Goyal *et al*. 2018**

307 In this study,[24] 21 patients with inflammatory bowel disease (any of Crohn's disease,
308 ulcerative colitis, or indeterminate colitis) received FMT and were available for follow-up.
309 We used remission at 30 days as the outcome. Each patient received stool from a
310 different donor, precluding any test of donor superiority. However, 16S rRNA
311 sequencing was performed, allowing for an analysis of donation characteristic
312 superiority, similar to Jacob *et al.* above. Similar to that study, there was no evidence for
313 an association between patient outcomes and the α-diversity (Figure 2a; $p = 0.97$,
314 Mann-Whitney *U* test) nor separation in donor microbiota (Figure 2b; $p = 0.7$,
315 PERMANOVA).

316

317 **Kump *et al*. 2017**

318 In this study,[23] 17 ulcerative colitis patients received 5 FMTs from 1 of 14 donors. 12
319 donors were used in 1 patient, 1 donor was used in 2 patients, and 1 donor was used in
320 3 patients. As in Goyal *et al*., this distribution of patients across donors does not permit

321 an assessment of donor superiority. However, like Jacob *et al.* and Goyal *et al*., this
322 study tested whether increased α-diversity was associated with improved patient
323 outcomes.
324
325 We detected no difference in diversity by patient outcome (Figure 3a; $p = 0.39$, Kruskal-
326 Wallis test; $p > 0.05$ for all 3 comparisons, Mann-Whitney U test), nor did we detect
327 separation in donor microbiota according to associated patient outcomes (Figure 3b; $p =$
328 0.15, PERMANOVA).
329
330 The original publication reported statistically significant differences between the α-
331 diversity of the 16 available donor stool samples associated with the 4 patients who
332 achieved remission compared to the diversity of 12 available samples associated with
333 the 4 patients who did not respond to FMT. Our analysis has a different conclusion,
334 finding no support for a donor effect, principally because we analyzed the data as only 8
335 independent data points, that is, one data point for each patient (4 remission, 4 no
336 response). By contrast, Kump *et al.* analyzed the data as if there were 28 independent
337 data points, that is, one data point for each donation (12 associated with remission, 16
338 associated with no response). Using donations rather than patients to determine N is a
339 statistical error known as "pseudoreplication".[41] Repeated measurements of the material
340 delivered to a patient do not constitute independent measures of the patient's outcome.
341 In other words, when testing for an association between patient outcomes and a
342 donation characteristic like α-diversity, the weight of the evidence is determined by the
343 number of patient outcomes, not by the number of microbiome measurements.[42]

Vermeire *et al*. 2016

In this study,[28] 14 patients with inflammatory bowel disease (either of 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, that patients with successful outcomes received more bacterially diverse donations (Mann-Whitney *U* test, $p = 0.012$). However, the underlying bacterial sequencing data was not available for re-analysis.

Nishida *et al*. 2017

In this study,[22] 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 7 donors whose patients met the primary endpoint, compared to 19 donors whose patients did not ($p = 0.69$). The study also tested whether the abundance of 10 taxa were differentially abundant among those two groups of donors, finding *p*-values less than 0.05 for 2 of 9 taxa, neither of which is statistically significant after Benjamini-Hochberg multiple hypothesis correction. The underlying bacterial sequencing data was not available for re-analysis.

DISCUSSION

In this study, we re-evaluated 9 studies that used FMT for IBD. In 7 of 9 cases, there was no statistically significant evidence of a donor effect. In one of the remaining

367 studies (Jacob *et al*.), there was a statistically significant difference in efficacy by pool of
368 donor stool, but not by donor, nor by the α-diversity of the stool or by donor microbiota
369 composition. We were unable to re-analyze the other study (Vermeire *et al.*) that
370 reported a statistically significant effect.
371
372 The major strengths of this study were the rigorous statistical approach used, which
373 avoided pseudoreplication and misinterpretation of results from an adaptive trial design,
374 and the uniformity of the re-analysis of the 16S rRNA microbiota composition data.
375
376 However, this study is also subject to two key limitations. First, our analysis considered
377 only donor effects and made no reference to patient factors like disease history. Given
378 the small sample sizes, adding patient factors as predictors would only decrease our
379 ability to detect donor effects, which was contrary to our main aim. We do not mean to
380 imply that all IBD patients are the same or should receive the same kind of treatment. In
381 contrast, a principal conclusion from this study is that patient factors, not donor factors,
382 are the main determinants of an FMT patient's clinical outcome.
383
384 Second, our analysis considered only dichotomous patient outcomes and a small
385 number of predictors, namely donors' identities (e.g., donor A vs. donor B), pools'
386 identities (pool A vs. pool B), 16S rRNA microbiota composition α-diversity, and
387 microbiota composition β-diversity. Previous work suggests that more proximal patient
388 biomarkers, like mucosal immune markers, are more likely to be able to detect donor

389 effects.[25] We caution, however, that attempts to find associations between every
390 conceivable donor factor, patient factor, and patient outcome risks "*p*-hacking".[43]
391
392 Overall, we found only very weak evidence for donor effects. On the other hand, the
393 small size of these studies —relative to the number of patients required to reliably
394 detect a plausible donor effect[25]— means that the data cannot definitively rule out a
395 large donor effect. It therefore remains undetermined if differences between donors are
396 clinically relevant to FMT for IBD. We propose that larger studies, with careful biomarker
397 selection and transparent reporting of results, is the best direction forward to
398 characterizing the donor effect for FMT in IBD.
399
400
401 **DECLARATIONS**
402
403 **Acknowledgements**
404 Sudarshan Paramsothy, Nadeem Kaakoush, and Rotem Sadovsky for assistance in
405 collecting the Paramsothy *et al.* data and for helpful conversations. Sam Costello for
406 assistance in collecting the Costello *et al.* data and for helpful conversations. Eric J.
407 Alm, Shrish Budree, Claire Duvallet, Justin O'Sullivan, Pratik Panchal, Marina Santiago,
408 Mark B. Smith, and Duane Wesemann for helpful conversations.
409
410 **Authorship statement**

SWO is the guarantor. SWO conceived the study, obtained data, performed the analysis, and wrote the manuscript. YG obtained data; improved the analysis, interpretation, and presentation of the results; and revised the manuscript. All authors approved the final version of the manuscript.

Statement of interests

SWO is employed by OpenBiome. YG is employed by and owns stock options in Finch Therapeutics. No specific funding was received for this work.

420 **REFERENCES**

421 1. Marchesi, J. R. *et al.* The gut microbiota and host health: a new clinical frontier.
422 *Gut* **65**, 330–339 (2016).
423 2. Allegretti, J. R., Mullish, B. H., Kelly, C. & Fischer, M. The evolution of the use of
424 faecal microbiota transplantation and emerging therapeutic indications. *Lancet*
425 *(London, England)* **394**, 420–431 (2019).
426 3. McDonald, L. C. *et al.* Clinical Practice Guidelines for Clostridium difficile Infection
427 in Adults and Children: 2017 Update by the Infectious Diseases Society of
428 America (IDSA) and Society for Healthcare Epidemiology of America (SHEA).
429 *Clin. Infect. Dis.* **66**, e1–e48 (2018).
430 4. Debast, S. B., Bauer, M. P. & Kuijper, E. J. European Society of Clinical
431 Microbiology and Infectious Diseases: Update of the Treatment Guidance
432 Document for Clostridium difficile Infection. *Clin. Microbiol. Infect.* **20**, 1–26
433 (2014).
434 5. Olesen, S. W., Panchal, P., Chen, J., Budree, S. & Osman, M. Global disparities
435 in faecal microbiota transplantation research. *Lancet Gastroenterol. Hepatol.* **5**,
436 241 (2020).
437 6. Lopetuso, L. R. *et al.* Fecal transplantation for ulcerative colitis: current evidence
438 and future applications. *Expert Opin. Biol. Ther.* **20**, 343–351 (2020).
439 7. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome.
440 *Nature* **505**, 559–563 (2014).
441 8. Turnbaugh, P. J. *et al.* The human microbiome project. *Nature* **449**, 804–10
442 (2007).

443 9. Rossen, N. G. *et al.* Findings From a Randomized Controlled Trial of Fecal
444 Transplantation for Patients With Ulcerative Colitis. *Gastroenterology* **149**, 110-
445 118.e4 (2015).
446 10. Wilson, B. C., Vatanen, T., Cutfield, W. S. & O'Sullivan, J. M. The Super-Donor
447 Phenomenon in Fecal Microbiota Transplantation. *Front. Cell. Infect. Microbiol.*
448 (2019) doi:10.3389/fcimb.2019.00002.
449 11. Olesen, S. W., Leier, M. M., Alm, E. J. & Kahn, S. A. Searching for superstool:
450 Maximizing the therapeutic potential of FMT. *Nature Reviews Gastroenterology*
451 *and Hepatology* (2018) doi:10.1038/s41575-018-0019-4.
452 12. Olesen, S. W., Gurry, T. & Alm, E. J. Designing fecal microbiota transplant trials
453 that account for differences in donor stool efficacy. *Stat. Methods Med. Res.*
454 (2018) doi:10.1177/0962280216688502.
455 13. Duvallet, C. *et al.* Framework for rational donor selection in fecal microbiota
456 transplant clinical trials. *PLoS One* **14**, e0222881 (2019).
457 14. Barnes, D. *et al.* Competitively Selected Donor Fecal Microbiota Transplantation.
458 *J. Pediatr. Gastroenterol. Nutr.* **67**, 185–187 (2018).
459 15. Osman, M., Abend, A., Panchal, P., Kassam, Z. & Budree, S. Does the Donor
460 Matter? Microbiome Sequencing to Evaluate Lower Donor Efficacy in Fecal
461 Microbiota Transplantation for Recurrent Clostridium Difficile Infection.
462 *Gastroenterology* **154**, S-25-S-26 (2018).
463 16. Ray, A. & Jones, C. Does the donor matter? Donor vs patient effects in the
464 outcome of a next-generation microbiota-based drug trial for recurrent Clostridium
465 difficile infection. *Future Microbiol.* **11**, 611–616 (2016).

466 17. Zuo, T. *et al.* Gut fungal dysbiosis correlates with reduced efficacy of fecal
467 microbiota transplantation in Clostridium difficile infection. *Nat. Commun.* **9**, 3663
468 (2018).
469 18. Moayyedi, P. *et al.* Fecal Microbiota Transplantation Induces Remission in
470 Patients With Active Ulcerative Colitis in a Randomized Controlled Trial.
471 *Gastroenterology* **149**, 102-109.e6 (2015).
472 19. Paramsothy, S. *et al.* Multidonor intensive faecal microbiota transplantation for
473 active ulcerative colitis: a randomised placebo-controlled trial. *Lancet* **389**, 1218–
474 1228 (2017).
475 20. Jacob, V. *et al.* Single Delivery of High-Diversity Fecal Microbiota Preparation by
476 Colonoscopy Is Safe and Effective in Increasing Microbial Diversity in Active
477 Ulcerative Colitis. *Inflamm. Bowel Dis.* **23**, 903–911 (2017).
478 21. Paramsothy, S. *et al.* Faecal Microbiota Transplantation for Inflammatory Bowel
479 Disease: A Systematic Review and Meta-analysis. *J. Crohn's Colitis* **11**, 1180–
480 1199 (2017).
481 22. Nishida, A. *et al.* Efficacy and safety of single fecal microbiota transplantation for
482 Japanese patients with mild to moderately active ulcerative colitis. *J.*
483 *Gastroenterol.* **52**, 476–482 (2017).
484 23. Kump, P. *et al.* The taxonomic composition of the donor intestinal microbiota is a
485 major factor influencing the efficacy of faecal microbiota transplantation in therapy
486 refractory ulcerative colitis. *Aliment. Pharmacol. Ther.* **47**, 67–77 (2018).
487 24. Goyal, A. *et al.* Safety, Clinical Response, and Microbiome Findings Following
488 Fecal Microbiota Transplant in Children With Inflammatory Bowel Disease.

489 *Inflamm. Bowel Dis.* **24**, 410–421 (2018).
490 25. Olesen, S. W. Power calculations for detecting differences in efficacy of fecal
491 microbiota donors. *medRxiv* (2020) doi:10.1101/2020.04.16.20068361.
492 26. Wasserstein, R. L., Schirm, A. L. & Lazar, N. A. Moving to a World Beyond " p <
493 0.05". *Am. Stat.* **73**, 1–19 (2019).
494 27. Budree, S. *et al.* The Association of Stool Donor Diet on Microbial Profile and
495 Clinical Outcomes of Fecal Microbiota Transplantation in Clostridium Difficile
496 Infection. *Gastroenterology* **152**, S630–S631 (2017).
497 28. Vermeire, S. *et al.* Donor Species Richness Determines Faecal Microbiota
498 Transplantation Success in Inflammatory Bowel Disease. *J. Crohn's Colitis* **10**,
499 387–394 (2016).
500 29. Kazerouni, A. & Wein, L. M. Exploring the Efficacy of Pooled Stools in Fecal
501 Microbiota Transplantation for Microbiota-Associated Chronic Diseases. *PLoS*
502 *One* **12**, e0163956 (2017).
503 30. Khanna, S. *et al.* Changes in microbial ecology after fecal microbiota
504 transplantation for recurrent C. difficile infection affected by underlying
505 inflammatory bowel disease. *Microbiome* **5**, 55 (2017).
506 31. R Core Team. *R: A language and environment for statistical computing*. (R
507 Foundation for Statistical Computing, 2019).
508 32. Fay, M. P. Confidence intervals that match Fisher's exact or Blaker's exact tests.
509 *Biostatistics* **11**, 373–374 (2010).
510 33. Mehta, C. R. & Hilton, J. F. Exact power of conditional and unconditional tests:
511 Going beyond the 2x2 contingency table. *Am. Stat.* (1993)

512 doi:10.1080/00031305.1993.10475946.
513 34. Anderson, M. J. A new method for non-parametric multivariate analysis of
514 variance. *Austral Ecol.* **26**, 32–46 (2001).
515 35. Oksanen, J. *et al.* vegan: Community Ecology Package. https://cran.r-
516 project.org/package=vegan (2019).
517 36. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects
518 models using lme4. *J. Stat. Softw.* (2015) doi:10.18637/jss.v067.i01.
519 37. Costello, S. P. *et al.* Effect of Fecal Microbiota Transplantation on 8-Week
520 Remission in Patients With Ulcerative Colitis. *JAMA* **321**, 156 (2019).
521 38. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome
522 data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
523 39. Amir, A. *et al.* Deblur Rapidly Resolves Single-Nucleotide Community Sequence
524 Patterns. *mSystems* **2**, (2017).
525 40. Bhatt, D. L. & Mehta, C. Adaptive Designs for Clinical Trials. *N. Engl. J. Med.* **375**,
526 65–74 (2016).
527 41. Hurlbert, S. H. Pseudoreplication and the Design of Ecological Field Experiments.
528 *Ecol. Monogr.* **54**, 187–211 (1984).
529 42. Lazic, S. E., Clarke-Williams, C. J. & Munafò, M. R. What exactly is ‘N’ in cell
530 culture and animal experiments? *PLOS Biol.* **16**, e2005282 (2018).
531 43. Head, M. L., Holman, L., Lanfear, R., Kahn, A. T. & Jennions, M. D. The Extent
532 and Consequences of P-Hacking in Science. *PLOS Biol.* **13**, e1002106 (2015).
533
534

FIGURES AND TABLES

537 **Table 1**: Summary of re-analyses.

Data source	No. patients	No. 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	—	—	—
Moayyedi *et al.*	38	6	No	—	—	—
Paramsothy *et al.*	78	14	No	No (0 p.p., 95% CI 0 to 46)	—	—
Costello *et al.*	38	19	No	No (0 p.p., 95% CI 0 to 74)	—	—
Jacob *et al.*	20	4	Yes (p = 0.02)	No (88 p.p., 95% CI 0 to 100)	No	No
Pool meta-analysis[‡]	137	36	No	No (41 p.p., 95% CI 0 to 88)	—	—
Goyal *et al.*	21	21	—	—	No	No
Kump *et al.*	17	14	—	—	No	No
Vermeire *et al.*	14	14	—	—	Reported	—
Nishida *et al.*	41	41	—	—	No	—

538 p.p.: Percentage points
539
540 a: Tests of donor superiority and pool superiority were using Fisher-Freeman-Halton or Fisher's exact test. Quantifications of pool effect are
541 differences in clinical efficacy between a very high quality pool (90th percentile) and low quality pool (10th percentile).
542 b: Using Mann-Whitney *U* test or Kruskal-Wallis test
543 c: Using PERMANOVA on Bray-Curtis β-diversity
544

545 **Table 2**: Tests of donor superiority, by donor, for Rossen *et al.* Donor labels are

546 arbitrary. “Success” means the patient reached the primary endpoint; “failure” means

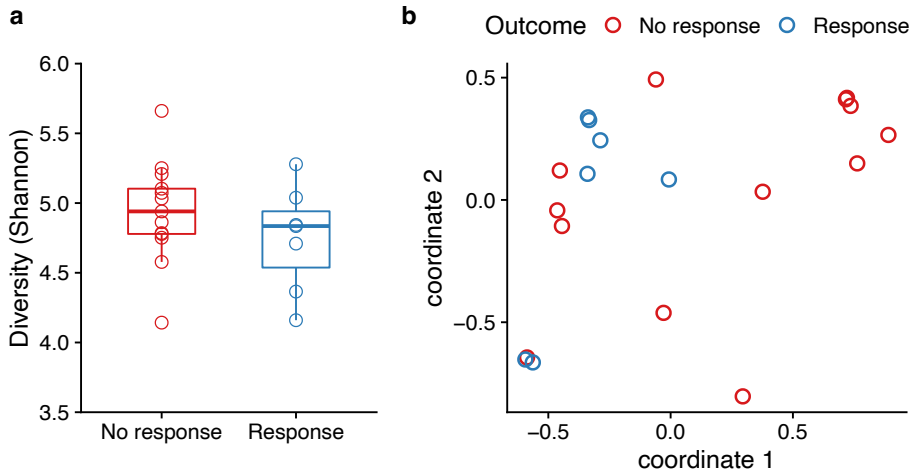
547 they did not.

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

548

549

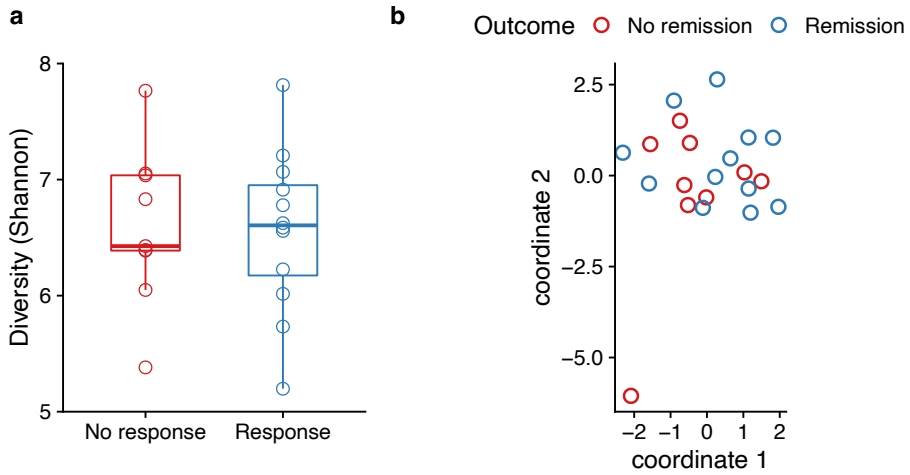
550 **Figure 1**: Microbiome analyses for Jacob *et al.* **a** Bacterial α-diversity of FMT pools do
551 not significantly differ by patient outcomes (p = 0.39, Mann-Whitney U test). Each point
552 represents a single patient. **b** Donor microbiota compositions (points; x- and y-
553 coordinates show multidimensional scaling using Bray-Curtis dissimilarity) do not
554 statistically significantly separate by associated patient outcome (color; p = 0.076,
555 PERMANOVA). Each point represents the composition of the pool used to treat an
556 individual patient.



557

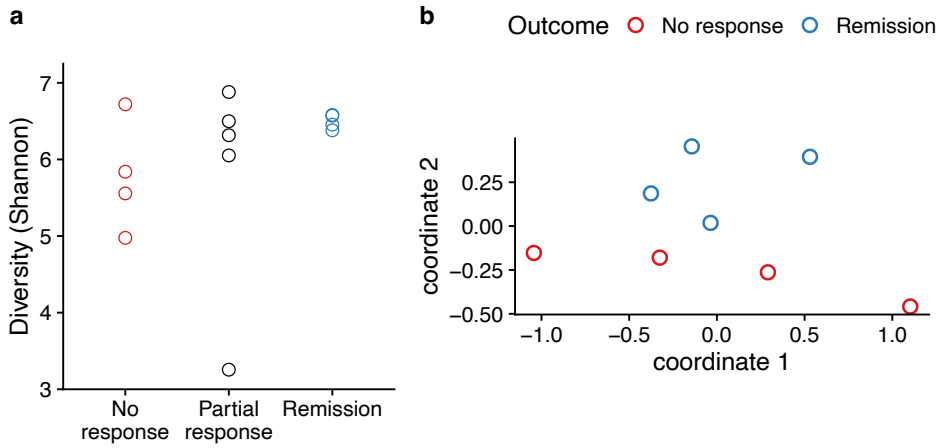
558

559 **Figure 2.** Microbiome analyses for Goyal *et al.* **a** Bacterial α-diversity of FMT pools do
560 not significantly differ by patient outcomes (p = 0.97, Mann-Whitney U test). Each point
561 represents a single patient. **b** Donor microbiota compositions (points; x- and y-
562 coordinates show multidimensional scaling using Bray-Curtis dissimilarity) do not
563 statistically significantly separate by associated patient outcome (color; p = 0.70,
564 PERMANOVA). Each point represents the composition of the first sample from the
565 donor used to treat an individual patient.



566

567 **Figure 3**: Microbiome analyses for Kump *et al.* **a** Bacterial α-diversity do not differ by
568 patient outcome (p = 0.39, Kruskal-Wallis test; p > 0.05 for all 3 comparisons, Mann-
569 Whitney U test). Each point represents a single patient. The diversity shown for each
570 patient (vertical axis) is the average of the diversity of the sampled FMTs used in that
571 patient. **b** Donor microbiota do not statistically significantly separate by associated
572 patient outcome (color; p = 0.15, PERMANOVA). Each point represents the composition
573 of the first donor sample administered to each patient.



574