Designing fecal microbiota transplant trials that account for differences in donor stool efficacy

Scott W. Olesen[1,⋆], Thomas Gurry[1,2,⋆], and Eric J. Alm[1,2,3,†]

[1] Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

[2] Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology, Cambridge, MA, USA

[3] The Broad Institute of MIT and Harvard, Cambridge, MA, USA

[⋆] These authors contributed equally to this work

[†] Corresponding author. Eric Alm, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. ejalm@mit.edu

1 Abstract

Fecal microbiota transplantation (FMT) is a highly effective intervention for patients suffering from recurrent *Clostridium difficile*, a common hospital-acquired infection. FMT's success as a therapy for *C. difficile* has inspired interest in performing clinical trials that experiment with FMT as a therapy for other conditions like inflammatory bowel disease, obesity, diabetes, and Parkinson's disease. Results from clinical trials that use FMT to treat inflammatory bowel disease suggest that, for at least one condition beyond *C. difficile*, most FMT donors produce stool that is not efficacious. The optimal strategies for identifying and using efficacious donors have not been investigated. We therefore examined the optimal Bayesian response-adaptive strategy for allocating patients to donors and formulated a computationally-tractable myopic heuristic. This heuristic computes the probability that a donor is efficacious by updating prior expectations about the efficacy of FMT, the placebo rate, and the fraction of donors that produce efficacious stool. In simulations designed to mimic a recent FMT clinical trial, for which traditional

power calculations predict ∼100% statistical power, we found that accounting for differences in donor stool efficacy reduced the predicted statistical power to ∼9%. For these simulations, using the heuristic Bayesian allocation strategy more than quadrupled the statistical power to ∼39%. We use the results of similar simulations to make recommendations about the number of patients, number of donors, and choice of clinical endpoint that clinical trials should use to optimize their ability to detect if FMT is effective for treating a condition.

2 Introduction

Fecal microbiota transplant (FMT), the transfer of stool from a healthy person into an ill person's gut, is a highly effective treatment for recurrent *Clostridium difficile* infections, which kill 30,000 Americans a year. Despite FMT's efficacy and increasingly widespread use, the biological mechanism by which FMT cures the infection is not fully understood [1, 2, 3, 4, 5, 6, 7]. FMT's success in treating *C. difficile* has generated interest in experiments that use FMT to treat other conditions related to the gut and the gut-associated microbiota [8, 9]. However, emerging evidence suggests using FMT for these other diseases will be more challenging. Notably, in a recent study by Moayyedi *et al.* [10] that used FMT to treat ulcerative colitis, patients appeared to respond to stool from only one of the six stool donors. Stool from all other donors was no more efficacious than placebo. These results suggest that the effectiveness of FMT can depend strongly on the choice of stool donor.

Ideally, information collected before or during a clinical trial could be used to identify which donors are efficacious. Predictions about a donor's efficacy could be repeatedly updated depending on that donor's performance, the performance of other donors, and prior expectations about donor efficacy and heterogeneity. There is, however, not enough clinical information about this variability or biological information about the mechanism by which FMT treats disease to create or validate a model for any particular indication.

We therefore present a general model of differences in donor stool efficacy that is not specific to a clinical indication or biomarker. Using this model, we describe the optimal, adaptive, Bayesian strategy for allocating donors and then formulate a computationally-tractable myopic Bayesian allocation heuristic.

Using simulations of clinical trials, we show that differences in donor efficacy and the trial's strategy for allocating donors can have substantial impacts on the trial's statistical power. We compare the performance of non-adaptive approaches for allocating patients to donors against a variation of a previously-studied adaptive algorithm (a "play the winner" strategy [11, 12, 13]) and our own Bayesian heuristic. We find that, in many cases, traditional non-adaptive allocation strategies are likely to falsely conclude that FMT in inefficacious. Adaptive strategies, however, can substantially increase a trial's ability to detect if FMT is efficacious.

3 Model of differences in donor stool efficacy

The results of the trial reported in Moayyedi *et al.* [10] raise the possibility that only some donors produce stool that is efficacious for treating ulcerative colitis, a type of inflammatory bowel disease. This possibility—that donors produce stool that is either efficacious or not—is a special case of a more general model of donor-patient interaction.

In this general model, a patient in the treatment arm can respond to the treatment under one of three scenarios:

1. the patient failed to respond to the FMT itself but did respond to the placebo effect,

2. the patient responded to the FMT and would have responded to the placebo effect if they had been given only a placebo, or

3. the patient responded to the FMT and would *not* have responded to a placebo.

We account for these three possibilities by treating the two effects—an effect from stool

and the placebo effect—as independent. Specifically, we say that stool has some "active ingredient" that, in the absence of a placebo effect, could cause a patient to respond.

In theory, stool from different donors could have different ingredient efficacies for different patients. Assuming that there are discrete classes of patients and donors, we write the ingredient efficacy of stool from a donor in class d for a patient in class t as $p_{\text{ingr}}^{(t,d)}$. The probability $p_{\text{eff}}^{(t,d)}$ that a patient in class t will respond to treatment (i.e., the treatment efficacy) using a donor in class d depends on $p_{\text{ingr}}^{(t,d)}$ and the probability p_{pl} that that patient will respond to the placebo effect:

$$1 - p_{\text{eff}}^{(t,d)} = \left(1 - p_{\text{ingr}}^{(t,d)}\right)(1 - p_{\text{pl}}). \tag{1}$$

The "matching" between patients and donors, if it exists, is not well understood and cannot yet be predicted. Furthermore, to minimize the patients' exposure to unknown pathogens, current FMT trials usually treat each patient with stool from only one donor. Thus, each patient is selected randomly from the different patient classes, so the observed treatment efficacy $p_{\text{eff}}^{(d)}$ is actually the expected value of $p_{\text{eff}}^{(t,d)}$ over the different patient classes t and patient placebo rates p_{pl}. Let $f_{\text{patient}}^{(t)}$ be the fraction of patients that are in class t and f_{pl} be the probability distribution function of the placebo rates so that:

$$1 - \mathbb{E}\left[p_{\text{eff}}^{(t,d)}\right]_t = \sum_t \int f_{\text{patient}}^{(t)} \left(1 - p_{\text{ingr}}^{(t,d)}\right)(1 - p_{\text{pl}})\, f_{\text{pl}}(p_{\text{pl}})\, \mathrm{d}p_{\text{pl}} \tag{2}$$

$$= \left(1 - \mathbb{E}\left[p_{\text{ingr}}^{(t,d)}\right]_t\right)\left(1 - \mathbb{E}\left[p_{\text{pl}}\right]\right). \tag{3}$$

Due to this limitation, we will model only the variables $p_{\text{eff}}^{(d)}$, $p_{\text{ingr}}^{(d)}$ and p_{pl} with the understanding that these variables are actually the mean values over distributions of patient characteristics.

The study by Moayyedi *et al.* [10] suggests that the majority of donors produce stool that has no efficacy beyond placebo, that is, that there is a class of donors for which $p_{\text{ingr}}^{(d)} = 0$. In light of the paucity of the available data about donor qualities, we made a

further simplification to the model: we assert that a fraction f_{eff} of donors have ingredient efficacy $p_{\text{ingr}} > 0$ and the remaining donors have $p_{\text{ingr}} = 0$ (i.e., $p_{\text{eff}} = p_{\text{pl}}$). This final model is summarized in Figure 1.

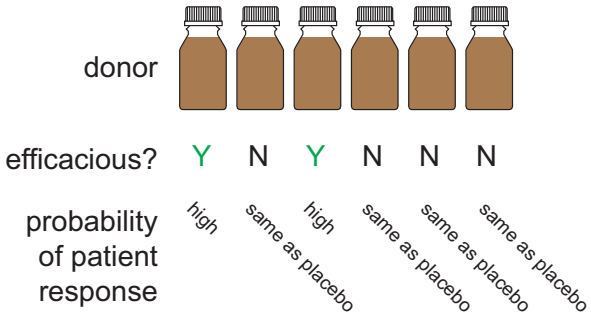

Figure 1: The model of differences in donor efficacy. In the model, donors are efficacious or not. Patients respond to FMT from an efficacious stool donor with probability p_{eff}. An FMT from an inefficacious stool donor is considered identical to a placebo, i.e., patients respond with probability p_{pl}. The fraction of donors in the general donor population that are efficacious is f_{eff}.

3.1 Bayesian model formulation

Adaptive donor assignment strategies aim to use the information derived from the patients' outcomes—and possibly some *a priori* beliefs about the values of the model parameters—to make decisions about how to assign donors. The donor assignment problem, then, is amenable to a Bayesian treatment, which will model the three parameters described above as well as the "qualities" $\boldsymbol{q} \in \{0,1\}^D$ of the D donors, where $q_i = 0$ indicates that donor i produces inefficacious (i.e., placebo-equivalent) stool and $q_i = 1$ indicates that that donor produces efficacious stool.

As described in the Results, we found that the results of simulations using this model were mostly insensitive to the choice of prior distribution on the parameters $(p_{\text{pl}}, p_{\text{ingr}}, f_{\text{eff}})$, so in most cases we used a simple uniform prior on these values. In one set of simulations,

Clinical trial result	Parameter	Point estimate	A	B
2 of 37 in placebo arm responded	p_{pl}	$2/37 = 0.054$	2	35
1 of 6 donors appeared efficacious	f_{eff}	$1/6 = 0.17$	1	5
7 of 18 patients allocated to the efficacious donor responded	p_{eff}	$7/18 = 0.39$	7	11

Table 1: Point estimates and hyperparameters for model parameters using clinical data [10]. A and B are hyperparameters for a beta distribution.

we articulated these priors using beta distributions:

$$P_{\text{pl,ingr}}(p_{\text{pl}}, p_{\text{ingr}}) = \text{Beta}(p_{\text{pl}}; A_{\text{pl}}, B_{\text{pl}}) \times \text{Beta}(p_{\text{eff}}; A_{\text{peff}}, B_{\text{peff}}) \tag{4}$$

$$P_{\text{feff}}(f_{\text{eff}}) = \text{Beta}(f_{\text{eff}}; A_{\text{feff}}, B_{\text{feff}}), \tag{5}$$

where the hyperparameters $A_{(\cdot)}$ and $B_{(\cdot)}$ can be drawn from prior clinical trial results. We chose to model the prior on p_{ingr} via a beta distribution on p_{eff} because p_{eff} is directly observable in terms of patient outcomes, while p_{ingr} can only be back-computed from p_{eff} and p_{pl} via equation (1). The clinical trial results for Moayyedi *et al.* [10] and their translation into the model parameters and prior hyperparameters are shown in Table 1.

We model the prior on the donor qualities $\boldsymbol{q}$ as a function of f_{eff}, the fraction of donors in the general population that are efficacious:

$$P(\boldsymbol{q}|f_{\text{eff}}) = \prod_{i=1}^{D} \begin{cases} f_{\text{eff}} & \text{if } q_i = 1 \\ 1 - f_{\text{eff}} & \text{if } q_i = 0 \end{cases} \tag{6}$$

The data used to update the priors are the numbers of patients treated successfully and unsuccessfully for each donor. The likelihood of the data $\boldsymbol{X}$ given these parameters is

$$P(\boldsymbol{X}|p_{\text{pl}}, p_{\text{ingr}}, f_{\text{eff}}, \boldsymbol{q}) = \prod_{i=1}^{D} \begin{cases} \text{Bin}(s_i, n_i; p_{\text{eff}}) & \text{if } q_i = 1 \\ \text{Bin}(s_i, n_i; p_{\text{pl}}) & \text{if } q_i = 0 \end{cases}, \tag{7}$$

where $\text{Bin}(s, n; p)$ is the binomial probability mass function with s successes among n

total trials with success probability p, s_i is the number of patients successfully treated with stool from donor i, and n_i is the total number of patients treated with stool from donor i.

Combining equations (6) and (7) shows that the posterior probability of the parameters $\boldsymbol{q}$ and $\boldsymbol{\pi} \equiv (p_{\text{pl}}, p_{\text{ingr}}, f_{\text{eff}})$ is

$$P(\boldsymbol{q}, \boldsymbol{\pi}|\boldsymbol{X}) \propto P(\boldsymbol{\pi}) \times \prod_{i=1}^{D} \begin{cases} f_{\text{eff}} \operatorname{Bin}(s_i, n_i; p_{\text{eff}}) & \text{if } q_i = 1 \\ (1 - f_{\text{eff}}) \operatorname{Bin}(s_i, n_i; p_{\text{pl}}) & \text{if } q_i = 0, \end{cases} \tag{8}$$

where $P(\boldsymbol{\pi})$ is the prior. The derivation of the posterior predictive probability of a patient responding to treatment with stool from a given donor is in the Appendix.

We found no closed-form solution to the integrals required to compute the predictive posterior probabilities, which we instead evaluated using numerical methods. Specifically, we used Monte Carlo integration with Suave (SUbregion-Adaptive VEgas), an importance sampling method combined with a globally adaptive subdivision strategy. Sampling for this integral was performed with Sobol pseudo-random numbers. The integrator was implemented in C++ using the Cuba package [14]. Wrappers for the integration routine were implemented in Python 3 and simulations were then parallelized to run on multiple cores to optimize computational run time [15].

3.2 Optimal donor allocation

Given the priors on the model parameters and the observed patient outcomes, what is the optimal choice for the next donor? This question is similar to other "bandit" problems [16]. Answering it requires evaluating the ramifications that a donor choice will have for the next patient as well as for the trial as a whole.

Let a trial *state* be a trial with some set of observed patient outcomes. Given a donor choice, each trial state has two *child* states: one in which the patient assigned to that donor responded to treatment, one in which the patient did not respond. The optimal

donor choice is the one with maximum expected utility U, where the utility reflects the outcome of the entire trial. The utility of a donor choice is

$$p\,U(S) + (1-p)\,U(F), \tag{9}$$

where p is the probability of a patient responding to treatment from that donor, S is the child state with a success from that donor, and F is the child state with a new failed patient.

FMT trials are often designed with a patient horizon (i.e., they will terminate after N patients have been treated), and it is natural to associate utilities with the terminal trial states. For example, it would be sensible to take the actions that are most likely to lead to a trial with a significant p-value as measured by frequentist statistics, so every terminal node with a significant p-value would have utility 1 and the others would have utility 0.

Equation (9) shows that the optimal donor choice at a non-terminal trial state depends only on the utility of its child nodes and each donor's probability of a successful treatment given that trial state. Thus, the optimal choice at an intermediate trial state can be found by backward induction:

1. Associate each terminal trial state (i.e., those with N patients treated) with a utility.

2. For each trial state with $N-1$ patients treated, compute the expected utility of each donor choice. Associate that state with the expected utility of the donor that maximizes that value.

3. Iterate backwards until the intermediate state has been reached. Select the donor with the maximum expected utility.

An algorithm for counting the number of unique trial states as a function of the number N of patients treated, derived in the Appendix, shows that a moderate-size trial with $N = 30$ has more than 10 million terminal trial states. Unfortunately, evaluating

the posterior predictive probability that a patient will respond to treatment from a donor is computationally intensive, making this optimal approach infeasible.

3.3 Deterministic myopic allocation heuristic

In light of the intractability of the optimal solution, we formulated a myopic heuristic that, rather than optimizing the expected utility of the entire trial, simply assigns the next patient to the donor that maximizes the probability that that patient will have a successful outcome. At each step in the trial, the predictive posterior probability of a successful patient outcome is computed for each donor, and the donor with the maximum probability is used. (In the case of a tie, a donor is chosen at random.)

4 Methods

4.1 Comparison of donor allocation strategies

To evaluate the effectiveness of our myopic heuristic, we performed simulations of clinical trials using four different allocation strategies. The first two strategies, block allocation and random allocation, are non-adaptive. The other two strategies, a randomized urn-based strategy and the myopic Bayesian heuristic, are response-adaptive.

Block allocation In a block allocation, patients are evenly allocated to donors. For example, if there are 30 patients and 6 donors, the first 5 patients are treated with the first donor, the second 5 patients are treated with the second donor, etc. (In clinical practice, patients would be randomized within the blocks.)

Random allocation In a random allocation, patients are allocated to donors at random. (On average, random allocations are similar to block allocations, so these two types of non-adaptive simulation yield similar results.)

Urn-based allocation Our urn-based allocation strategy is a variation on the "play-the-winner" strategies designed and studied as an ethical [17] and statistically-rigorous way to decide how to allocate patients to a treatment arm when a trial includes more than one treatment arm [18, 19, 20]. In this study, we used the generalized Pólya's urn [21] with parameters $w = 1$, $\alpha = 3$, $\beta = 0$, and without replacing the drawn ball.

Myopic Bayesian heuristic This is the strategy described in the previous section. Except when noted, a uniform prior on the three parameters $(p_{\mathrm{pl}}, p_{\mathrm{ingr}}, f_{\mathrm{eff}})$ was used.

4.2 Simulated clinical trials

The expected fraction of patients allocated to efficacious donors and the statistical power of clinical trials using different donor allocation strategies were estimated using simulated clinical trials. In each simulation, the three model parameters (p_{pl}, p_{eff}, f_{eff}) and the number of patients in the trial were fixed. For each combination of the trial parameters, 10,000 lists of 6 donors each were randomly generated. Donors were designated as efficacious or not efficacious by random chance according to the frequency of efficacious donors f_{eff}. The same donor lists were used for simulations for each of the allocation strategies.

In one set of simulations, the number of donors was varied among 1, 3, 5, 10, 15, and 30. For those simulations, lists of 30 donors were generated for each parameter set and trial iteration. The lists were truncated for the simulations using less than 30 donors.

For each allocation strategy and donor list, a trial was simulated. In each simulation, a patient allocated to an efficacious donor responds to the treatment with probability p_{eff}. Patients allocated to inefficacious donors or to the placebo arm respond with probability p_{pl}. For adaptive allocations, the outcomes from all the previous patients treatment were determined before the donor for the next patient was selected. An equal number of patients was allocated to the treatment and placebo arms. (This equal allocation to the control arm was intended to preserve statistical power, which would probably have

decreased if the myopic algorithm were allowed to allocate patients away from the placebo arm [22, 23].)

Clinically-relevant parameter values Simulations were performed for all combinations of parameter values selected to reflect clinically-relevant possibilities:

- The placebo rate p_{pl} is either 0.05 (a low placebo rate consistent with stringent, objective outcomes; e.g., endoscopic Mayo score [10, 24]) or 0.25 (a high placebo rate consistent with self-reported, subjective outcomes [25, 26]).
- The efficacy p_{eff} of efficacious donors is either 0.4 (similar to the value in Table 1) or 0.95 (efficacy of FMT to treat *C. difficile* infection).
- The frequency f_{eff} of efficacious donors is either 0.15 (similar to the value in Table 1) or 0.9 (reflective of the fact that almost any well-screened donor produces stool that can successfully treat *C. difficile* infection).
- The number of patients in each of the treatment and control arms is 15, 30, or 60, corresponding to a range of patient numbers typical for Phase I and small Phase II clinical trials.

Of these combinations, the set of values most similar to the one in Table 1 is $p_{\text{pl}} = 0.05$, $p_{\text{eff}} = 0.4$, $f_{\text{eff}} = 0.15$, $N_{\text{patients}} = 30$.

Computing statistical power After determining the outcome of all the patients in the trial, the p-value of a one-sided Fisher's exact test (asserting that the response rate in the treatment arm was greater) was calculated. The proportion of simulations that produced $p < 0.05$ was the estimate of the statistical power for that allocation strategy under those trial parameters. Confidence intervals were calculated using the method of Clopper and Pearson [27]. Values are rounded to two or three significant digits.

5 Results

5.1 Trials using adaptive strategies allocate more patients to efficacious donors

The purpose of adaptive donor allocation strategies is to identify and use efficacious donors. We therefore expected that simulated trials that use adaptive strategies would allocate more patients to efficacious donors (compared to simulated trials that used the block or random donor allocation strategies).

For every parameter set simulated, the average fraction of patients allocated to efficacious donors was greater in the adaptive strategies (urn-based and myopic Bayesian) than in the non-adaptive strategies (block and random; Table S1). The two non-adaptive allocation strategies performed almost identically: for each parameter set, their results differed by less than 1 percentage point. The two adaptive strategies performed similarly: for half of the parameter sets, their results differed by less than 2 percentage points. In the remaining parameter sets, their results varied by between 2 and 9 percentage points.

When efficacious donors are common ($f_{\text{eff}} = 0.9$), the adaptive and non-adaptive strategies performed similarly (Fig. 2). In other cases, the performance of the two strategies differed substantially. For example, for the parameterization most similar to the one in Table 1, the random strategy allocated 15% of patients to efficacious donors while the myopic Bayesian strategy allocated 41% of patients to efficacious donors.

5.2 Trials using adaptive allocation have higher statistical power

Because trials that used the adaptive donor allocation strategies allocated more patients to efficacious donors than the trials that used the non-adaptive strategies, we expected that trials using adaptive strategies would have greater statistical power.

The adaptive strategies consistently yielded higher statistical powers than the non-adaptive strategies (Table 2). When efficacious donors are rare, the performance gap is

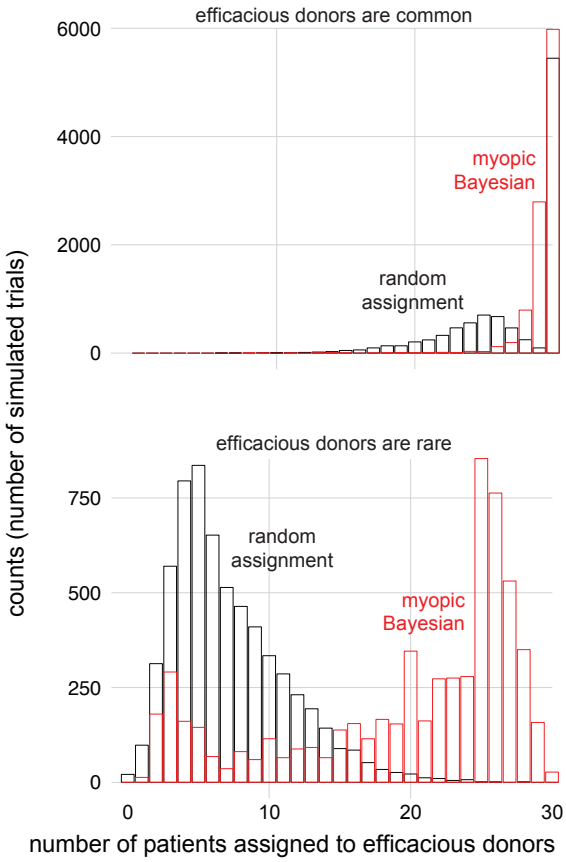

Figure 2: The Bayesian myopic strategy allocates more patients to efficacious donors. (top) For the parameterization most similar to the one in Table 1 (top), the Bayesian myopic heuristic (red) allocated more patients to efficacious donors than random allocation (black) did. When efficacious donors are common (bottom; same parameters as top but with $f_{\text{eff}} = 0.9$), the two strategies allocate similar numbers of patients to efficacious donors. For visual clarity, only trials in which there was at least one efficacious donor are shown.

larger. For example, for the parameterization most similar to the one in Table 1, a trial that uses random allocation is expected to have 9% power, while the myopic Bayesian strategy can deliver 39% power. The gap in performance is smallest when selecting a donor at random is likely to yield an efficacious donor: among the trials with $f_{\text{eff}} = 0.9$, the adaptive and non-adaptive statistical powers differed by less than 6 percentage points.

Low statistical powers when f_{eff} is small are likely due to the fact that if *all* the available donors are not efficacious, then no allocation strategy should make a trial achieve significance. For example, if only 15% of donors are efficacious ($f_{\text{eff}} = 0.15$) and there are only 6 donors (the number used in these simulations), then we expect that 38% of

Trial parameters				FMT power (%)				Naive power (%)
p_{pl}	p_{eff}	f_{eff}	N_{patients}	block	random	urn	Bayesian	
0.05	0.4	0.15	15	4.57	5.04	17.2	19.1	68.1
			30	8.44	8.89	33.7	39.4	93.8
			60	23.0	24.1	53.5	58.0	100
		0.9	15	59.1	59.7	63.4	64.3	67.8
			30	87.7	87.4	92.1	93.2	94.2
			60	99.2	99.1	99.8	99.9	100
	0.95	0.15	15	19.9	21.3	60.3	60.7	100
			30	33.9	34.2	61.4	61.6	100
			60	55.9	53.2	63.2	63.1	100
		0.9	15	99.8	99.5	100	100	100
			30	100	100	100	100	100
			60	100	100	100	100	100
0.25	0.4	0.15	15	2.80	2.88	3.43	3.77	11.4
			30	4.44	4.55	7.16	7.27	25.5
			60	6.55	6.15	13.6	13.9	47.3
		0.9	15	10.5	10.3	10.5	10.7	11.7
			30	21.6	21.3	23.0	23.0	25.0
			60	40.7	40.7	44.5	44.8	47.2
	0.95	0.15	15	9.04	10.3	38.6	47.4	99.4
			30	20.0	20.3	60.0	60.1	100
			60	32.6	32.7	63.4	62.7	100
		0.9	15	94.3	93.9	98.2	99.0	99.6
			30	99.6	99.4	100	100	100
			60	100	100	100	100	100

Table 2: Adaptive strategies yield clinical trials with higher statistical power. "FMT power" is the power computed by simulating the results of trials that would occur if the frequency of efficacious donors if f_{eff}. "Naive power" is the power computed in the situation in which all donors are efficacious (i.e., $f_{\text{eff}} = 1.0$). All 95% confidence intervals on these values are within 1 percentage point of the reported value and are not shown.

trials will have no good donors (using the binomial distribution function). We therefore separately analyzed the simulated trials in which no donors were efficacious and the simulated trials in which at least one donor was efficacious (Table S2). When no donors are efficacious, trials with adaptive or non-adaptive strategies have ∼0% power. Among trials with at least one efficacious donor, the difference in statistical power between adaptive and non-adaptive strategies is greater than the difference computed using the results of all trials.

Conversely, the power computed in traditional calculations that do not account for

differences in donor efficacy (i.e., that assume that all donors are efficacious, or equivalently $f_{\text{eff}} = 1.0$) is, in many cases, substantially higher than the power computed when accounting for differences in donor efficacy (Table 2, column "Naive powers"). For example, for the parameter set most similar to the one in Table 1, the naive calculation predicts 94% power, but the calculation that accounts for differences in donor efficacy predicts only 9% power for non-adaptive allocation strategies. The differences between the powers computed by the naive method and our approach is largest when f_{eff} is small.

5.3 Performance of adaptive strategies depend on their parameterization

In these simulations, we varied the actual values of p_{pl} and p_{eff} but we always initialized the adaptive algorithms the same ways. To determine the sensitivity of the adaptive allocation algorithms' performance to their initialization, we simulated trials in which the actual model parameters were fixed but the algorithms' initializations varied (Table S3 and Table S4). The myopic Bayesian algorithm's performance was mostly robust to the parameterization of its prior distribution except when the prior was strong and inaccurate. Accurate priors, weak priors, and uniform priors provide comparable performance. In contrast, the urn algorithm delivered widely varying powers, from 15% to 40%, depending on its parameterization.

5.4 Increasing the number of available donors benefits the adaptive Bayesian strategy

Increasing the number of available donors increases the probability that at least one of them will be efficacious. We therefore determined, for each donor allocation strategy, the number of available donors that optimized the trial's expected power. Simulations showed that increasing the size of the donor "pool" almost always increased the power of trials using the myopic Bayesian donor allocation but, depending on the parameter set,

could increase or decrease the power of trials using other allocation strategies (Table S5).

Table 3 shows how donor selection strategy and model parameter values affects the optimal number of donors. Notably, when efficacious donors are uncommon ($f_{\text{eff}} = 0.15$) and only moderately efficacious ($p_{\text{eff}} = 0.4$), the non-adaptive strategies perform optimally when only one donor is used. In other words, when using non-adaptive strategies in this parameter regime, it is wiser to take the 15% chance of picking a single efficacious donor than it is to distribute patients across many donors, allotting around 15% of them to efficacious donors. In contrast, the myopic Bayesian donor allocation almost always benefits from a larger donor pool.

Trial parameters			Optimal N_{donors}			
f_{eff}	p_{eff}	p_{pl}	random	block	urn	Bayesian
0.15	0.4	0.05	1	1	10	10–30
0.15	0.4	0.25	1	1	5	10
0.15	0.95	0.05	3–15	3–5	15–30	15–30
0.15	0.95	0.25	3	3	10	10–30
0.9	0.4	0.05	1–30	3–30	3–30	3–30
0.9	0.4	0.25	1–30	1–30	1–30	5–30
0.9	0.95	0.05	3–30	3–30	3–30	3–30
0.9	0.95	0.25	3–30	3–30	3–30	3–30

Table 3: The optimal number of donors varies by donor selection strategy and model parameter values. For each parameter value, trials were simulated using 1, 3, 5, 10, 15, and 30 donors. (The number of patients was fixed at 30.) The number of donors that optimized the expected power was identified. If multiple numbers of donors yielded powers within 0.05 of the optimal value, all those numbers are reported as a range.

6 Discussion

6.1 Model limitations

We chose to use a simple model for a simple use case because, in the absence of data about the treatment histories of hundreds of patients using dozens of donors, we do not believe that more complicated models will be more useful to aiding trial design. The model's greatest weakness is that it cannot be validated, but it is exactly the model's purpose

to improve the probability of collecting the kind of information that could validate or invalidate it. In light of the dearth of data, we developed a simple model, and it could be that the simplifications we made limit the model's validity. For example, the model assumes that each donor produces efficacious stool or inefficacious stool (when in fact there is probably day-to-day and donor-to-donor variation in stool efficacy) and that all patients receive one course of treatment (while, say, patients who do not respond to a first treatment might be treated with stool from a different donor).

In our formulation of the model, we only considered optimizing the assignment of patients to donors within the treatment arm, thus "protecting" the size of the placebo arm. Future work could determine if there is a benefit to adaptively assigning patients to the control or treatment arms based on estimations of the donors' quality. We also assumed in our simulations that the outcome from all the previous patients treatments are known before the donor for the next patient is selected. In reality, patients in an FMT trial overlap. The urn-based method can still be used for overlapping patients [21], but the myopic Bayesian method would require some modification. The urn-based approach is also randomized, which is desirable in clinical trials because it can reduce certain kinds of bias [11]. Future work could adapt the myopic Bayesian heuristic, which is deterministic, for adaptive randomization [28] and for delays between patient allocation and outcome observation.

6.2 Differences in donor efficacy should be accounted for in trial design

Our results entail recommendations to clinicians. First, the powers we computed here are, in many cases, well below the powers computed assuming that all donors are efficacious. We therefore encourage researchers to consult our predictions about statistical power when deciding on the size of their trials.

Second, a high placebo rate can substantially decrease the statistical power of an FMT

trial. We therefore encourage researchers to use the most stringent outcome measurement possible (e.g., an endoscopic Mayo score for inflammatory bowel disease).

Third, adaptive donor allocation strategies consistently delivered higher statistical power than traditional, non-adaptive approaches. We therefore recommend that researchers use such an adaptive strategy. The urn-based strategy has the advantages that similar response-adaptive strategies may be familiar to clinicians, it is randomized, and it is simple to implement. However, an urn-based strategy needs to be carefully parameterized: a badly-parameterized urn-based strategy performs similarly to random allocation. The adaptive Bayesian donor allocation algorithm performs well even when using the "default" settings (a uniform prior) but is complex and deterministic. To fully leverage this strategy, a clinician would need to consult the algorithm's output after every patient outcome and follow the algorithm's deterministic instructions, which might introduce bias.

Fourth, adaptive algorithms benefit from having access to a "bank" of 10 or more donors. Researchers hoping to achieve the full benefits of adaptive donor selection must be prepared to change donors multiple times during the trial.

Finally, researchers reporting about FMT trials should include information about the donors, notably how many donors were used and what proportion of patients allotted to each donor responded to treatment. This information will help future researchers account for differences in donor efficacy.

6.3 Future research may identify mechanistic explanations of FMT's efficacy

The adaptive allocation strategies we described here have a narrow aim: to increase the number of successful patient outcomes in a trial. In theory, an adaptive trial design is capable of more. For example, if it were hypothesized that FMT succeeded or failed because of the presence or absence of some particular microbial species in the donor's

stool, then an adaptive trial design could recommend donor choices that aim to identify that critical species.

We did not pursue a hypothesis-centric approach because we believe it is premature. Even the mechanism by which FMT treats *C. difficile*, the most well-studied case, remains unclear. We expect that strong hypotheses about mechanism will come from retroactive comparison of efficacious vs. inefficacious stool *after* clinical trials have definitively show that FMT is effective for treating some disease. Our study aims to do exactly this. Until then, we hope that our results about adaptive donor allocation help more patients benefit from FMT and will help clinicians identify those conditions that FMT can treat.

7 Acknowledgements

This material is based on work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1122374 and by the Center for Microbiome Informatics and Therapeutics at MIT.

References

[1] Stuart H. Cohen, Dale N. Gerding, Stuart Johnson, Ciarán P. Kelly, Vivian G. Loo, L. Clifford McDonald, Jacques Pepin, and Mark H. Wilcox. Clinical practice guidelines for *Clostridium difficile* infection in adults. *Infect Control Hosp Epidemiol*, 31(5):431–455, 2010.

[2] Els van Nood, Anne Vrieze, Max Nieuwdorp, Susana Fuentes, Erwin G. Zoetendal, Willem M. de Vos, Caroline E. Visser, Ed J. Kuijper, Joep F. W. M. Bartelsman, Jan G. P. Tijssen, Peter Speelman, Marcel G. W. Dijkgraaf, and Josbert J. Keller. Duodenal infusion of donor feces for recurrent *Clostridium difficile*. *N Engl J Med*, 368(5):407–415, 2013.

[3] Ciarán P. Kelly and J. Thomas LaMont. *Clostridium difficile* – more difficult than ever. *N Engl J Med*, 359(18):1932–1940, 2008.

[4] Neha Alang and Colleen R. Kelly. Weight gain after fecal microbiota transplantation. *Open Forum Infect Dis*, 2(1), 2015.

[5] Simone S. Li, Ana Zhu, Vladimir Benes, Paul I. Costea, Rajna Hercog, Falk Hildebrand, Jaime Huerta-Cepas, Max Nieuwdorp, Jarkko Salojärvi, Anita Y. Voigt, Georg Zeller, Shinichi Sunagawa, Willem M. de Vos, and Peer Bork. Durable coexistence of donor and recipient strains after fecal microbiota transplantation. *Science*, 352(6285):586–589, 2016.

[6] Zain Kassam, Christine H. Lee, Yuhong Yuan, and Richard H. Hunt. Fecal microbiota transplantation for *Clostridium difficile* infection: Systematic review and meta-analysis. *Am J Gastroenterol*, 108(4):500–508, 2013.

[7] A. Khoruts and M. J. Sadowsky. Understanding the mechanisms of faecal microbiota transplantation. *Nat Rev Gastroenterol Hepatol*, 2016.

[8] Loek P. Smits, Kristien E. C. Bouter, Willem M. de Vos, Thomas J. Borody, and Max Nieuwdorp. Therapeutic potential of fecal microbiota transplantation. *Gastroenterology*, 145(5):946–953, 2013.

[9] Michael J. Sadowsky and Alexander Khoruts. Faecal microbiota transplantation is promising but not a panacea. *Nat Microbiol*, 1, 2016.

[10] Paul Moayyedi, Michael G. Surette, Peter T. Kim, Josie Libertucci, Melanie Wolfe, Catherine Onischi, David Armstrong, John K. Marshall, Zain Kassam, Walter Reinisch, and Christine H. Lee. Fecal microbiota transplantation induces remission in patients with active ulcerative colitis in a randomized controlled trial. *Gastroenterology*, 149(1):102–109.e6, 2015.

[11] Thomas D. Cook and David L. DeMets. *Introduction to Statistical Methods for Clinical Trials*. Chapman & Hall/CRC, 2008.

[12] Robert H. Bartlett, Dietrich W. Roloff, Richard G. Cornell, Alice French Andrews, Peter W. Dillon, and Joseph B. Zwischenberger. Extracorporeal circulation in neonatal respiratory failure: A prospective randomized study. *Pediatrics*, 76(4):479–487, 1985.

[13] Shein-Chung Chow and Mark Chang. *Adaptive Methods in Clinical Trials*. Chapman & Hall/CRC, 2007.

[14] Thomas Hahn. Cuba – a library for multidimensional numerical integration. *Comput Phys Commun*, 168(2):78–95, 2005.

[15] Ole Tange. Gnu parallel – the command-line power tool. *;login: The USENIX Magazine*, 36(1):42–47, 2011.

[16] Donald A. Berry and Bert Fristedt. *Bandit Problems: Sequential Allocation of Experiments*. Springer Netherlands, 1985.

[17] James H. M. Hung, Robert T. O'Neill, Sue-Jane Wang, and John Lawrence. A regulatory view of adaptive/flexible clinical trial design. *Biom J*, 48(4):565–573, 2006.

[18] L. J. Wei, R. T. Smythe, D. Y. Lin, and T. S. Park. Statistical inference with data-dependent treatment allocation rules. *J Am Stat Assoc*, 85(409):156–162, 1990.

[19] L. J. Wei and S. Durham. The randomized play-the-winner rule in medical trials. *J Am Stat Assoc*, 73(364):840–843, 1978.

[20] M. Zelen. Play the winner rule and the controlled clinical trial. *J Am Stat Assoc*, 64(325):131–146, 1969.

[21] L. J. Wei. The generalized polya's urn design for sequential medical trials. *Ann Stat*, 7(2):291–296, 1979.

[22] L. Trippa, E. Q. Lee, P. Y. Wen, T. T. Batchelor, T. Cloughesy, G. Parmigiani, and B. M. Alexander. Bayesian adaptive randomized trial design for patients with recurrent glioblastoma. *J Clin Oncol*, 30(26):3258–3263, 2012.

[23] Sofía S. Villar, James Wason, and Jack Bowden. Response-adaptive randomization for multi-arm clinical trials using the forward looking gittins index rule. *Biometrics*, 71(4):969–978, 2015.

[24] Noortje G. Rossen, Susana Fuentes, Mirjam J. van der Spek, Jan G. Tijssen, Jorn H.A. Hartman, Ann Duflou, Mark Löwenberg, Gijs R. van den Brink, Elisabeth M.H. Mathus-Vliegen, Willem M. de Vos, Erwin G. Zoetendal, Geert R. D'Haens, and Cyriel Y. Ponsioen. Findings from a randomized controlled trial of fecal transplantation for patients with ulcerative colitis. *Gastroenterology*, 149(1):110–118, 2015.

[25] Chinyu Su, Gary R. Lichtenstein, Karen Krok, Colleen M. Brensinger, and James D. Lewis. A meta-analysis of the placebo rates of remission and response in clinical trials of active crohn's disease. *Gastroenterology*, 126(5):1257–1269, 2004.

[26] Chinyu Su, James D. Lewis, Brittany Goldberg, Colleen Brensinger, and Gary R. Lichtenstein. A meta-analysis of the placebo rates of remission and response in clinical trials of active ulcerative colitis. *Gastroenterology*, 132(2):516–526, 2007.

[27] C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934.

[28] Y. Cheng and D. A. Berry. Optimal adaptive randomized designs for clinical trials. *Biometrika*, 94(3):673–689, 2007.

A Mathematical appendix

A.1 Posterior probability of patient response

Let S_i be the event that the next patient responds to treatment with stool from donor i. The posterior predictive probability of S_i is

$$P(S_i|\boldsymbol{X}) = \int_{\boldsymbol{\pi}} \sum_{\boldsymbol{q}} P(S_i|\boldsymbol{\pi}, \boldsymbol{q}) \underbrace{P(\boldsymbol{\pi}, \boldsymbol{q}|\boldsymbol{X})}_{\text{posterior}} \, \mathrm{d}\boldsymbol{\pi}. \tag{10}$$

The probability $P(S_i|\boldsymbol{\pi}, \boldsymbol{q})$ is simply p_{eff} is $q_i = 1$ and p_{pl} if $q_i = 0$. Thus,

$$\begin{aligned} P(S_i|\boldsymbol{X}) \propto \int_{\boldsymbol{\pi}} & \underbrace{[p_{\text{eff}} f_{\text{eff}} \operatorname{Bin}(s_i; n_i, p_{\text{eff}}) + p_{\text{pl}}(1 - f_{\text{eff}}) \operatorname{Bin}(s_i, n_i; p_{\text{pl}})]}_{\text{term for donor } i} \\ & \times \prod_{j \neq i} \underbrace{[f_{\text{eff}} \operatorname{Bin}(s_j, n_j; p_{\text{eff}}) + (1 - f_{\text{eff}}) \operatorname{Bin}(s_j, n_j; p_{\text{pl}})]}_{\text{terms for other donors}} \times \underbrace{P(\boldsymbol{\pi})}_{\text{prior}} \, \mathrm{d}\boldsymbol{\pi} \end{aligned} \tag{11}$$

The constant of proportionality is the same for all i: the quantity in equation (11) should be divided by

$$\int_{\boldsymbol{\pi}} \prod_{i=1}^{D} [f_{\text{eff}} \operatorname{Bin}(s_i, n_i; p_{\text{eff}}) + (1 - f_{\text{eff}}) \operatorname{Bin}(s_i, n_i; p_{\text{pl}})] \, P(\boldsymbol{\pi}) \, \mathrm{d}\boldsymbol{\pi}. \tag{12}$$

Because the binomials in the equations (11) and (12) have the same constants, $\operatorname{Bin}(s, n; p)$ can be replaced by $p^s(1-p)^f$ in these equations.

A.2 Counting trial states

How many unique trial states are there for a given number of patients N?

Naively, a trial state can be specified by a string that specifies with donor was used and what outcome was observed (e.g., $AsAfBs$ means that donor A has a success, then a failure, then donor B had a success). A trial with N patients can have up to N donors, meaning that there are N^N possible chains of donors and 2^N possible chains of outcomes,

for $(2N)^N$ possible trial states.

However, many of those states are identical. In particular, we consider donors identical unless they have a different number of successful or failed patient outcomes (e.g., $AsAs$ is identical to $BsBs$), and we consider trial states identical if the same results are achieved in a different order (e.g., $AsAf$ is identical to $AfAs$). Thus, a unique trial set is a set of ordered triples (d_i, s_i, f_i), where d_i is the number of donors that each have had s_i successful and f_i failed patient outcomes.

Theorem 1. *Let $w(N, M)$ be the number of trial states with $N \geq 0$ total patients with at most $M \geq 0$ patients in any donor. $W(N, M)$ is defined by the recursion:*

$$w(0, M) = 1 \tag{13}$$

$$w(1, M) = 2 \tag{14}$$

$$w(N, 1) = N + 1 \tag{15}$$

$$w(N, M) = \sum_{i=0}^{\lfloor \frac{N}{M} \rfloor} \binom{M+i}{M} \times w(N - iM, M - 1) \tag{16}$$

The number of trial states with N patients is $w(N, N)$.

Proof. We then justify each of the recursion rules. For $N = 0$ (i.e., no patients), there is only one trial state: $\{(0, 0, 0)\}$. For $N = 1$ (i.e., one patient), there are two trial states: $\{(1, 1, 0)\}$ and $\{(1, 0, 1)\}$. For $M = 1$ (i.e., at most one patient per donor), there are $N + 1$ trial states: $\{(i, 1, 0), (N - i, 0, 1)\}$ for $0 \leq i \leq N$. In the general case, let i be the number of donors with M patients. This number can range from 0 to $\lfloor \frac{N}{M} \rfloor$. For each of those donors, there are $M + 1$ possible success/failure pairs: $(j, M - j)$ for $0 \leq j \leq M$. Thus, there are $\binom{M+i}{M}$ possible trial states for just those i donors. This leaves $N - iM$ patients to be distributed among the remaining donors, who can each have at most $M - 1$ patients. □

Some representative results are shown in Table S6.

B Supplementary Tables

p_{pl}	p_{eff}	f_{eff}	N_{patients}	block		random		urn		Bayesian	
				mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
0.05	0.4	0.15	15	0.15	0.15	0.15	0.17	0.29	0.30	0.32	0.33
			30	0.15	0.14	0.15	0.16	0.36	0.35	0.41	0.38
			60	0.15	0.14	0.15	0.15	0.44	0.39	0.49	0.42
		0.9	15	0.9	0.12	0.9	0.14	0.95	0.075	0.96	0.068
			30	0.9	0.12	0.9	0.13	0.97	0.051	0.98	0.041
			60	0.9	0.12	0.9	0.13	0.98	0.03	0.99	0.022
	0.95	0.15	15	0.15	0.15	0.15	0.17	0.44	0.36	0.51	0.42
			30	0.15	0.15	0.15	0.16	0.52	0.42	0.56	0.45
			60	0.15	0.15	0.15	0.15	0.57	0.45	0.6	0.47
		0.9	15	0.9	0.13	0.9	0.14	0.97	0.047	0.99	0.028
			30	0.9	0.12	0.9	0.13	0.98	0.027	1.0	0.013
			60	0.9	0.12	0.9	0.13	0.99	0.014	1.0	0.0069
0.25	0.4	0.15	15	0.15	0.15	0.15	0.17	0.20	0.25	0.21	0.29
			30	0.15	0.15	0.15	0.16	0.24	0.30	0.24	0.34
			60	0.15	0.15	0.15	0.15	0.29	0.34	0.28	0.38
		0.9	15	0.9	0.13	0.9	0.14	0.93	0.14	0.94	0.15
			30	0.9	0.12	0.9	0.13	0.94	0.13	0.95	0.14
			60	0.9	0.12	0.9	0.13	0.96	0.11	0.96	0.14
	0.95	0.15	15	0.15	0.15	0.15	0.17	0.36	0.34	0.45	0.42
			30	0.15	0.15	0.15	0.16	0.45	0.38	0.52	0.44
			60	0.15	0.15	0.15	0.15	0.53	0.42	0.57	0.46
		0.9	15	0.9	0.13	0.9	0.14	0.96	0.076	0.99	0.056
			30	0.9	0.12	0.9	0.13	0.98	0.048	0.99	0.029
			60	0.9	0.12	0.9	0.13	0.99	0.029	1.0	0.016

Table S1: Fraction of patients allocated to efficacious donors. The same simulated trials were analyzed to create Table 2 and this table.

Parameters		Simulation conditions		Power (%)			
p_{eff}	f_{eff}	Donor pool quality	$N_{\text{simulations}}$	random	block	urn	Bayesian
0.4	0.15	all trials	10000	8.9	8.4	34	39
0.4	0.15	no efficacious donors	3794	0.37	0.29	0.4	0.16
0.4	0.15	some efficacious donors	6206	14	13	54	63
0.4	0.9	some efficacious donors	10000	87	88	92	93
0.95	0.15	all trials	10000	34	34	61	62
0.95	0.15	no efficacious donors	3851	0.62	0.44	0.34	0.31
0.95	0.15	some efficacious donors	6149	55	55	100	100
0.95	0.9	some efficacious donors	10000	100	100	100	100

Table S2: Power of simulated clinical trials, conditioned on presence of efficacious donors. A subset of the data used to create Table 2 (only those simulations with $N_{\text{patients}} = 30$ and $p_{\text{pl}} = 0.05$) was analyzed by separately estimating the power for the trials in which no donors were efficacious ("no efficacious donors") and in which at least one donor was efficacious ("some efficacious donors"). "All simulations" shows the same data as in Table 2. For $f_{\text{eff}} = 0.9$, none of the 10,000 simulations had a donor pool with no efficacious donors. All 95% confidence intervals are within 1 percentage point of the reported value and are not shown.

	Hyperparameter values						
Hyperparameterization	A_{placebo}	B_{placebo}	A_{peff}	B_{peff}	A_{feff}	B_{feff}	Power (%)
From Table 1	2	35	7	11	1	5	41.1
Weak, accurate	0.2	3.5	0.7	1.1	0.1	0.5	41.2
Strong, accurate	20	350	70	110	10	50	40.4
Weak, inaccurate	0.4	3.3	1.4	0.4	0.2	0.4	40.9
Strong, inaccurate	40	330	140	40	20	40	34.4
Uniform	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	39.6

Table S3: Sensitivity of simulated clinical trial power to parameterization of the myopic Bayesian algorithm parameters. Starting from the parameter described in Table 1, 10,000 trials using the myopic Bayesian donor allocation were simulated for each of six different parameterizations of the Bayesian algorithm. "Accurate" means that the prior is centered around the true value (i.e., that $\frac{A}{A+B}$ equals the true value); "inaccurate" means that the prior is centered at approximately double the true value (but that $A + B$ has been held constant). "Strong" means that every hyperparameter is ten-fold greater; "weak" means that every hyperparameter is ten-fold smaller. "Uniform" means a uniform prior was used for all parameters. All 95% confidence intervals are within 1 percentage point of the reported value and are not shown.

w	α	β	replace?	Power (%)
1	3	0	true	28.5
			false	35.6
		1	true	13.9
			false	14.2
	6	0	true	33.2
			false	41
		1	true	16.2
			false	16.4
3	3	0	true	19.7
			false	29.1
		1	true	13.2
			false	13.1
	6	0	true	24.8
			false	32.6
		1	true	15.6
			false	15.3

Table S4: Sensitivity of simulated clinical trial power to urn parameterization. Using the parameter set described in Table 1, 10,000 trials using the urn-based donor allocation were simulated for each of several combinations of the parameters for the urn model as described in [21]. We also vary whether drawn balls are replaced or not. All 95% confidence intervals are within 1 percentage point of the reported value and are not shown.

Parameters				Power (%)			
f_{eff}	p_{eff}	p_{pl}	N_{donors}	random	block	urn	Bayesian
0.15	0.4	0.05	1	15.5	15.3	15.5	14.8
			3	12.4	12.2	28.4	30.9
			5	9.87	9.67	33.8	38
			10	7.49	6.93	38.2	44.3
			15	6.91	6.38	34.6	45.9
			30	6.17	5.26	16.6	45.9
		0.25	1	6.08	6.15	5.89	5.58
			3	4.64	4.21	6.7	6.89
			5	4.65	4.7	7.28	7.1
			10	4.37	4.18	6.2	7.52
			15	4.48	4.09	5.41	7.07
			30	4.25	4.3	4.67	6.89
	0.95	0.05	1	15.2	15.1	15	14.9
			3	34.4	37.2	39	39.2
			5	35	36.1	56.3	56.5
			10	33.7	33.3	80.1	80.4
			15	33.5	32.6	91	90.6
			30	31.6	30.4	87.4	93.9
		0.25	1	17.1	17.3	17	16.8
			3	23.1	22.6	39	39.9
			5	21.3	20.9	55.2	55.5
			10	18.8	17.7	69.7	69.7
			15	17.4	16.8	65.4	69.9
			30	16.4	15.4	43.2	69.4
0.9	0.4	0.05	1	84.9	84.7	84.6	84.3
			3	85.9	85.8	91.8	92.4
			5	87.4	87.2	92.4	92.4
			10	88.4	88.9	92.3	92.8
			15	88	89	91.8	92.9
			30	89.1	89.3	91.1	92.9
		0.25	1	22	22.3	22.5	22.1
			3	21.8	21.2	23.1	22.7
			5	21.7	21.3	23.1	23.2
			10	21.7	21.3	22.7	23.7
			15	22.2	21.5	22.5	24
			30	21.9	21.6	22.1	23.5
	0.95	0.05	1	90.1	90.1	90.1	90.4
			3	99.5	99.7	99.9	99.9
			5	100	100	100	100
			10	100	100	100	100
			15	100	100	100	100
			30	100	100	100	100
		0.25	1	90.4	90.4	90.4	90.5
			3	97.7	98.1	99.9	99.9
			5	99.1	99.4	100	100
			10	99.8	99.9	100	100
			15	99.8	99.9	100	100
			30	99.9	100	100	100

Table S5: Dependence of simulated clinical trial power on number of available donors. For each parameter value, trials were simulated using 1, 3, 5, 10, 15, and 30 donors. The number of patients was fixed at 30. All 95% confidence intervals are within 1 percentage point of the reported value and are not shown.

# patients	# trial states
1	2
2	6
3	14
4	33
5	70
6	149
7	298
8	591
9	1,132
10	2,139
15	39,894
20	5.58×10^{5}
25	6.39×10^{6}
30	6.29×10^{7}
40	4.34×10^{9}
50	2.14×10^{11}
60	8.22×10^{12}

Table S6: The number of possible trial states for a given number of patients, assuming that the number of available donors is equal to the number of patients.