Supplemental Information for "The public health impact of fecal microbiota transplantation"

L. Mamoon and S. W. Olesen

1 Model formulation

Model of disease course The model is summarized in Figure 1. In the model, there are X individuals at risk of a primary *C. difficile* infection. Over a time period $1/\alpha$, a proportion p_0 of the at-risk individuals experience a primary CDI episode. (In other words, the population X suffers infections at a rate αp_0.) Call this population of individuals suffering a primary episode I_1.

Patients suffering a primary episode I_1 suffer second episodes, transitioning to population I_2 at a rate αp_1. The fraction $(1 - p_1)$ of patients who do not suffer a second episode are cured and return to the at-risk pool, or they die, have colectomies, or are otherwise removed from pool of patients at risk for a second episode. We assume that the size of the at risk population X remains constant, so that any deaths or other removals from the system are replaced by incoming susceptible patients. (In the mathematical model, the arrow connecting, say, I_1 back to X encodes both cured individuals as well as individuals who are removed from the populations of interest and are replaced by new at-risk individuals.)

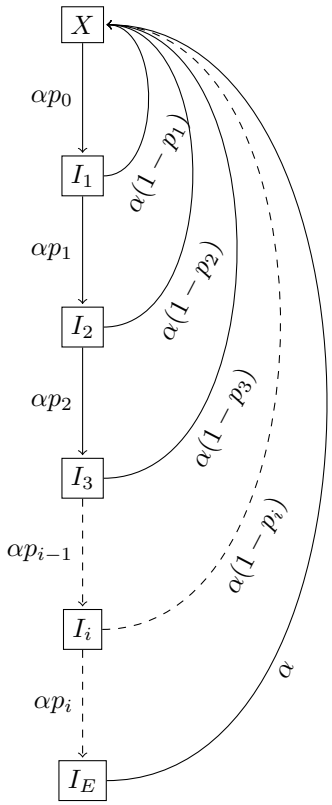
Similarly define the transitions from I_2 to I_3, and so forth. We assume that, after some maximum number E of episodes, patients do not recur, either because they die, have a colectomy, or are otherwise removed from the populations of interest.

The effect of FMT Relative to standard therapy, FMT reduces the probability of a further episode. We assume that FMT has the same, independent benefit in each episode. Consistent with a previous model [1], we also assume that multiply recurrent patients (i.e., those on their third or later episode) may be treated with FMT.

Let the *coverage* be the proportion of multiply recurrent patients treated with FMT instead of standard antibiotic therapy. Thus, the recurrence probability p_i for third or later episodes is a mixture of the recurrence probability p_{abx} on standard therapy and the probability p_{FMT} on FMT:

$$p_i = (1 - u)p_{\text{abx}} + u\, p_{\text{FMT}} \text{ for } i \geq 3,$$

where $0 \leq u \leq 1$ is the coverage.



Supplemental Figure 1: Model formulation. Dashed lines stand for the chain of populations $3 < i < E$.

Equilibrium analysis We already asserted that at-risk individuals are replaced so that the population X remains constant:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = 0$$

If the recurrence probabilities p_i are constant in time, then the populations will equilibrate:

$$0 = \frac{\mathrm{d}I_1}{\mathrm{d}t} = \alpha p_0 X - \alpha\left[(1-p_1)+p_1\right] I_1 \implies I_1 = p_0 X$$

Similarly, $I_2 = p_1 I_1$, $I_3 = p_2 I_2$, and so forth.

Over a period of time τ, there are $\tau\alpha p_0 X = \tau\alpha I_1$ new primary CDI cases. Similarly, there are $\tau\alpha p_1 I_1 = \tau\alpha I_2$ new secondary episodes, $\tau\alpha I_3$ new tertiary episodes, and so forth. Define the annual number of i-th episode cases as $C_i = \alpha I_i \times (1 \text{ year})$. These case counts follow the same recurrence relations: $C_2 = p_1 C_1$, $C_3 = p_2 C_2$, etc.

The total number of CDI episodes is:

$$\begin{aligned} C_\bullet = \sum_{i=1}^{E} C_i &= C_1 + C_2 + C_3 + C_4 + \ldots + C_E \\ &= C_1\left(1 + p_1 + p_1 p_2 + p_1 p_2 p_3 + \ldots + p_1 \cdots p_{E-1}\right) \end{aligned}$$

2 Outcomes

Proportion of episodes that are multiply recurrent ($f_{\geq 3}$) Given our assumption that the p_i are identical for $i \geq 3$, the summation at and after the third episode simplifies:

$$\begin{aligned} C_{\geq 3} &= C_3 + C_4 + C_5 + \ldots \\ &= C_3\left(1 + p_3 + p_3 p_4 + \ldots\right) \\ &= C_3 \sum_{i=0}^{E-3} p_3^i \\ &= C_3 \frac{1 - p_3^{E-2}}{1 - p_3} \end{aligned}$$

Note that $C_3 = C_1 p_1 p_2$. The total number of episodes therefore simplifies to:

$$C_\bullet = C_1\left(1 + p_1 + p_1 p_2 \frac{1 - p_3^{E-2}}{1 - p_3}\right)$$

And thus, the fraction of all CDI episodes that are multiply-recurrent is independent of the total number of cases $C_\bullet$:

$$\begin{aligned} f_{\geq 3} = \frac{C_{\geq 3}}{C_\bullet} &= \frac{C_{\geq 3}}{C_1 + C_2 + C_{\geq 3}} \\ &= 1 \Big/ \left(\frac{C_1 + C_2}{C_{\geq 3}} + 1 \right) \\ &= 1 \Big/ \left(\frac{1 + p_1}{p_1 p_2 \frac{1 - p_3^{E-2}}{1 - p_3}} + 1 \right) \end{aligned}$$

Number of FMTs (N_{FMT}) The total number of FMTs is $u\, C_{\geq 3} = u f_{\geq 3} C_\bullet$. Note, however, that $f_{\geq 3}$ is a function of p_3, which is itself a function of u.

Number needed to treat (NNT) The number needed to treat is, as u increases, the ratio of the marginal increase in the number of FMTs to the marginal decrease in the number of total episodes:

$$\text{NNT} = -\frac{\mathrm{d}(u\, C_{\geq 3})}{\mathrm{d}(C_{\geq 3})} = -\frac{\frac{\mathrm{d}(u\, C_{\geq 3})}{\mathrm{d}u}}{\frac{\mathrm{d}C_{\geq 3}}{\mathrm{d}u}} = -\left(u + \frac{C_{\geq 3}}{\frac{\mathrm{d}C_{\geq 3}}{\mathrm{d}u}} \right)$$

Some algebra shows that:

$$\begin{aligned} \frac{\mathrm{d}C_{\geq 3}}{\mathrm{d}u} &= C_3 \frac{\mathrm{d}}{\mathrm{d}u} \left(\frac{1 - p_3^{E-2}}{1 - p_3} \right) \\ &= C_3 \left[\frac{-(E-2) p_3^{E-3} (1 - p_3) - (1 - p_3^{E-2})(-1)}{(1 - p_3)^2} \right] \frac{\mathrm{d}p_3}{\mathrm{d}u} \\ &= C_{\geq 3} \left[\frac{1}{1 - p_3} - (E-2) \frac{p_3^{E-3}}{1 - p_3^{E-2}} \right] (p_{\text{FMT}} - p_{\text{abx}}) \end{aligned}$$

Note that $p_{\text{FMT}} < p_{\text{abx}}$, so that:

$$\text{NNT} = \left\{ \left[\frac{1}{1 - p_3} - (E-2) \frac{p_3^{E-2}}{1 - p_3^{E-2}} \right] (p_{\text{abx}} - p_{\text{FMT}}) \right\}^{-1} - u$$

Note again that p_3 is itself a function of u. This dependence on u disappears when the maximum number of episodes E increases:

$$\lim_{E \to \infty} \text{NNT} = \frac{1 - p_{\text{abx}}}{p_{\text{abx}} - p_{\text{FMT}}}$$

3 Parameters

To develop confidence intervals, we bootstrap around the reported values of $C_\bullet$, p_1, and p_2.

Total number of cases ($C_\bullet$) $C_\bullet$ is estimated in Guh *et al.* [2] as a mean value μ with a 95% confidence interval. Assuming that $C_\bullet \sim N(\mu, \sigma)$, we can vary σ to best match the reported confidence intervals. Specifically, optimize σ to minimize the sum-of-squares deviation between the reported 95% confidence intervals and the ones computed using σ.

Probability of first recurrence (p_1) We use the estimates from Rajasingham *et al.* [1] of multiple parameters to develop a single estimate for p_1.

First, we model the proportions of disease that is nonsevere, severe, or fulminant with a Dirichlet distribution. We interpret the point estimates given in Rajasingham *et al.* as the means of the Dirichlet proportions and the given ranges as 95% confidence intervals. We then find a single value α_0 (sum of the Dirichlet concentration parameters) such that the Dirichlet distribution with those means best matches those 95% confidence intervals.

Second, for the remaining variables, we treat them as beta-distributed and search for values of α and β that best match the reported 95% confidence intervals.

Finally, we draw a point estimate and random variates for the proportion of patients who experience a first recurrence along any of these 5 paths:

1. Initial nonsevere disease, treatment with vancomycin, treatment failure
2. Initial nonsevere disease, treatment with vancomycin, initial cure, recurrence
3. Initial severe disease, treatment with vancomycin, treatment failure
4. Initial severe disease, treatment with vancomycin, initial cure, recurrence
5. Initial fulminant disease, treatment with vancomycin, initial cure, recurrence

We then fit a beta distribution to the variates. We draw from that beta distribution in the later simulations.

We elected to use vancomycin alone as the modeled treatment option for 2 reasons. First, using a mixture of metronidazole, vancomycin, and fidaxomicin would require introducing 6 parameters (2 for the proportions of each drug used; times 3 for primary episodes, first recurrence, and later recurrences) for which estimates are not readily available. Second, although the model is sensitive to p_{abx}, the estimates for recurrence probabilities for second and later recurrences in Rajasingham *et al*. are similar for vancomycin and fidaxomicin.

Probability of second recurrence (p_2) We use the estimates from Rajasingham *et al.* to estimate a single beta distribution according to one of two paths: fail, or cure and recur.

We repeat a similar method for the third and further recurrences for vancomycin and FMT.

Probability of recurrence when using antibiotics (p_{abx}) or FMT (p_{FMT}) These values are drawn from Rajasingham *et al.*, using vancomycin for p_{abx}.

Maximum number of episodes (E) This baseline value value was drawn from McFarland *et al.* [3].

4 Supplemental Figures and Tables

Parameter	Point estimate	Range	Distribution	Source
p_1	29%	25% to 33%	beta	[1]
p_2	41%	29% to 53%	beta	[1]
p_{abx}	52%	35% to 70%	beta	[1]
p_{FMT}	21%	17% to 25%	beta	[1]
$C_\bullet$	462,100	428,600 to 495,600	normal	[2]
E	15	10 to 20	uniform	[3]

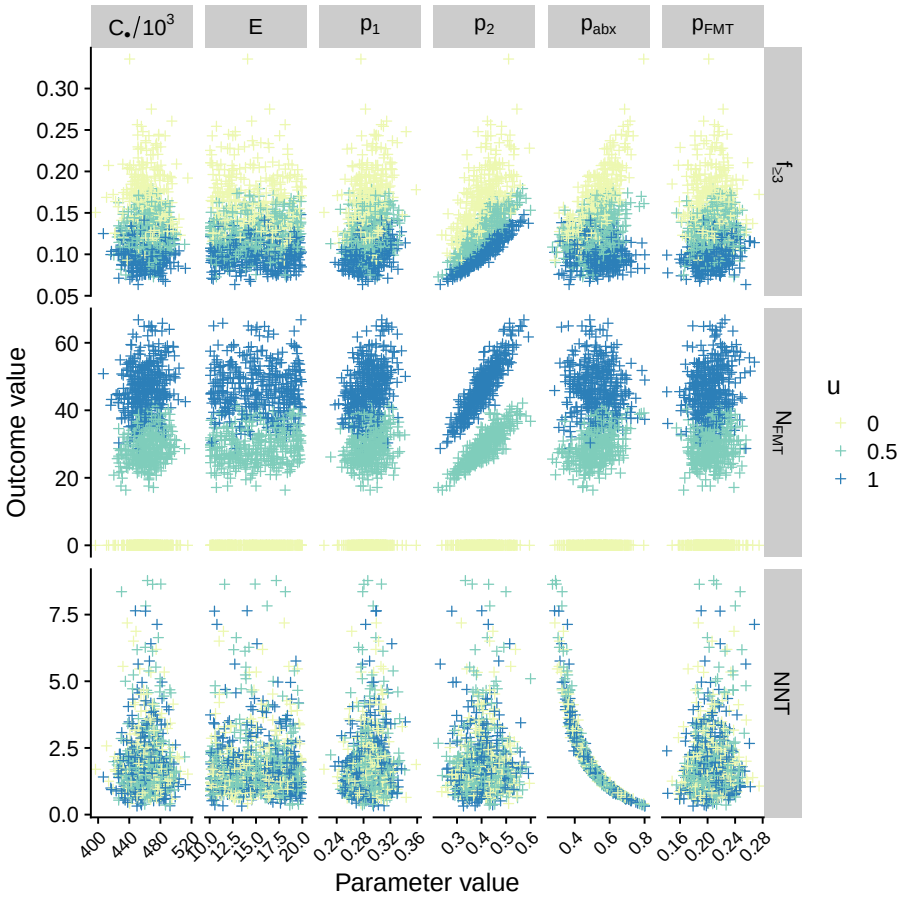
Supplemental Table 1: Model parameters. For beta- and normal-distributed variables, the reported range is the 95% confidence interval. For the uniform-distributed variable, it is the minimum and maximum values.

Outcome	Coverage (u)	Point estimate (95% CI)
N_{FMT}	0%	0
	50%	29,000 (20,000 to 39,000)
	100%	48,000 (34,000 to 62,000)
$f_{\geq 3}$	0%	16% (11% to 24%)
	50%	13% (9% to 17%)
	100%	10% (7% to 13%)
NNT	0%	1.5 (0.6 to 4.7)
	50%	1.5 (0.6 to 4.7)
	100%	1.5 (0.6 to 4.7)

Supplemental Table 2: Model outcomes.

Parameter	Outcome	Sensitivity (%)
p_{abx}	NNT	-28
u	$f_{\geq 3}$	-2.2
p_{FMT}	N_{FMT}, $f_{\geq 3}$	1.4
p_{abx}	N_{FMT}, $f_{\geq 3}$	3.6
p_{FMT}	NNT	6.4
p_1	N_{FMT}, $f_{\geq 3}$	6.8
u	N_{FMT}	7.8
p_2	N_{FMT}, $f_{\geq 3}$	8.7
$C_\bullet$	N_{FMT}	10

Supplemental Table 3: Sensitivity analysis. "Sensitivity" is the increase in the outcome per 10% increase in the parameter. Only combinations of parameters and outcomes with sensitivities greater than 0.1% are shown.



Supplemental Figure 2: Input parameter values (horizontal axis, top labels) and output outcome values (vertical axis, right labels) for three values of FMT coverage u (colors). For visual clarity, the 0.3% of bootstrap replicates with a number needed to treat outside of 0 to 10 were excluded, and a random subset of 1,000 of those qualifying bootstraps is shown.

References

[1] Radha Rajasingham, Eva A Enns, Alexander Khoruts, and Byron P Vaughn. Cost-effectiveness of treatment regimens for *Clostridioides difficile* infection: An evaluation of the 2018 Infectious Diseases Society of America Guidelines. *Clinical Infectious Diseases*, April 2019.

[2] Alice Y. Guh, Yi Mu, Lisa G. Winston, Helen Johnston, Danyel Olson, Monica M. Farley, Lucy E. Wilson, Stacy M. Holzbauer, Erin C. Phipps, Ghinwa K. Dumyati, Zintars G. Beldavs, Marion A. Kainer, Maria Karlsson, Dale N. Gerding, and L. Clifford McDonald. Trends in U.S. burden of *Clostridioides difficile* infection and outcomes. *New England Journal of Medicine*, 382(14):1320–1330, April 2020.

[3] Lynne V. McFarland, Christina M. Surawicz, Moshe Rubin, Robert Fekety, Gary W. Elmer, and Richard N. Greenberg. Recurrent *Clostridium difficile* disease: Epidemiology and clinical characteristics. *Infection Control & Hospital Epidemiology*, 20(01):43–50, January 1999.