Probiotic responder identification in cross-over trials for constipation using a Bayesian statistical model considering lags between intake and effect periods

Shion Hosoda[1,2*], Yuichiro Nishimoto[3], Yohsuke Yamauchi[3], Takuji Yamada[3,4], and Michiaki Hamada[1,2,5*]

[1]Department of Electrical Engineering and Bioscience, Graduate School of Advanced Science and Engineering, Waseda University, Tokyo, Japan
[2]Computational Bio Big-Data Open Innovation Laboratory, AIST-Waseda University, Tokyo, Japan
[3]Metagen Inc., Yamagata, Japan
[4]Department of Life Science and Technology, Tokyo Institute of Technology, Tokyo, Japan
[5]Graduate School of Medicine, Nippon Medical School, Tokyo, Japan
[*]To whom correspondence should be addressed.

Abstract

Recent advances in microbiome research have led to the further development of microbial interventions, such as probiotics and prebiotics, which are potential treatments for constipation. However, the effects of probiotics vary from person to person; therefore, the effectiveness of probiotics needs to be verified for each individual. Individuals showing significant effects of the target probiotic are called responders. A statistical model for the evaluation of responders was proposed in a previous study. However, the previous model does not consider the lag between intake and effect periods of the probiotic. It is expected that the lag exists when probiotics are administered and when they are effective. In this study, we propose a Bayesian statistical model to estimate the probability that a subject is a responder, by considering the lag between intake and effect periods. In synthetic dataset experiments, the proposed model was found to outperform the base model, which did not factor in the lag. Further, we found that the proposed model could distinguish responders showing large uncertainty in terms of the lag between intake and effect periods.

1 Introduction

Recent advances in microbiome research have resulted in the rapid development of microbial interventions, such as probiotics and prebiotics, which are potential treatments for constipation [1]. Probiotics are living microbes that benefit the host when ingested in sufficient quantities and are reported to improve defecation frequencies and treat constipation [1, 2]. The effects of probiotics vary from person to person [3]. Individuals exhibiting significant effects of probiotics are called "responders" [4]; each responder exhibits a significant effect of a different probiotic. That is, different responders respond to different probiotics, and individual differences make it difficult to evaluate the effects of probiotics. Therefore, experimental designs used for probiotic research should consider the individual differences between subjects.

One type of sophisticated experimental design is a cross-over trial, in which each subject takes both the target probiotic and a placebo. A cross-over trial comprises the following steps: (1) Each individual is first administered a capsule containing the target probiotic or placebo for several days. (2) After a washout period, which lasts several weeks and is set to remove the effects of the former capsule, each individual is administered another capsule (containing either a probiotic or a placebo) for a specific period. Cross-over trials are widely applied in various fields of research, including research related to probiotics [5], prebiotics [6], neurorehabilitation [7], and spinal manipulation [8]. The main advantage of cross-over trials is that they enable the evaluation of individual differences [9], which are supposed to be determined by attributes such as gender, genetics, and habits. Accordingly, datasets obtained from a cross-over trial require a reasonable analysis method that considers individual differences.

An approach to estimate individual differences has already been conducted in a previous study. Nakamura *et al.* evaluated improvements in defecation frequencies using a Weibull regression model [10]. They revealed individual differences in the improvement of defecation frequency by grouping subjects into three groups: strong responders, weak responders, and non-responders. However, their model used the unreasonable assumption that the effects of the target probiotics start on the day the probiotic is administered to the subject. A previous study suggested that orally ingested material should be excreted for one or more days [11]. In addition, it has been estimated that it can take more than ten hours for microbes to increase dramatically [12] and at least two days for food to alter the gut microbiome [13]. Therefore, analyses that do not consider the lag between intake and effect periods can lead to a misidentification of responders, especially in short-term intervention experiments.

In this study, we propose a Bayesian statistical model for estimating the efficacy of a target probiotic in improving defecation frequency, considering the lag between intake and effect periods (Figure 1) and individual differences in a cross-over trial dataset. This study aimed to identify responders accurately using cross-over trial datasets where the lag between intake and effect periods exists. The proposed model is based on the segmented linear regression model, which represents each periodic term using linear regression, and has discrete parameters for lag days. The proposed model evaluates the cumulative sum of the number of times a subject defecates. An individual can be evaluated based on the posterior probability that the individual is a responder to probiotics. With the proposed model, we estimated whether each subject is a responder using synthetic datasets and the real dataset used in the previous study [10]. We compared the results of the proposed model with those of a base model that did not consider the lag period. Our analysis showed that considering the effect of time lag was useful in the synthetic dataset experiments. Real data experiments showed that the proposed model estimated the posterior distribution considering the effect lag and led to different conclusions from those of the base model. We found that the proposed model could eliminate uncertain responders (responders whose response to a probiotic is uncertain) according to the lag between intake and effect periods.

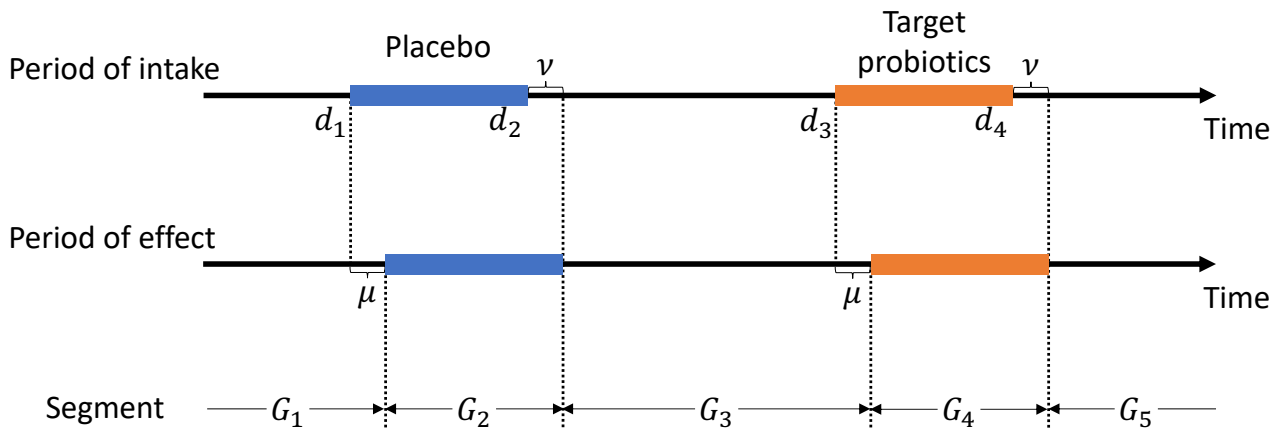

Figure 1: Schematic illustration of the effect of lag in cross-over trials. Blue and orange boxes represent the placebo and target probiotics periods, respectively. d_1, d_2, d_3, and d_4 represent the start day index of the first capsule, the end day index of the first capsule, the start day index of the second capsule, and the end day index of the second capsule, respectively. μ and ν are the lag of effect start and end, respectively. G_t shows the term of the day in the t-th segment. The time when an effect of the probiotic was observed was delayed compared to the time the probiotic was ingested. In this case, the subject was first administered the placebo capsule and then the target probiotic capsule.

2 Materials and methods

2.1 Overview

Here, we provide an overview of the proposed model. The proposed model requires a dataset represeting the cumulative sum of the number of defecation events collected from a cross-over trial. Figure 1 shows the schematic illustration of the respective cross-over trial whose dataset was used in this study. The terms of the trial are divided into several periodic terms with respect to capsule intake/effect and are denoted as "segments." Here, $S = 5$ is used in this figure and the dataset used in this study, where S is the number of segments. The proposed Bayesian statistical model is based on segmented linear models for the cumulative sum of the number of defecation events. The proposed model can consider the lag between intake and effect periods (*cf.* Section 1) using the number of lag days as discrete parameters.

2.2 Generative process

Here, we describe the generative process of the proposed model. The proposed model demands the cumulative sum of the defecation frequency $y_i (i = 1 \dots N)$, where N is the number of days in the entire trial term. This model is for one subject. Let $\alpha \in \mathbb{R}$, $\eta \in \mathbb{R}$, and $\tilde{\eta} \in \mathbb{R}$ be the logarithmic defecation frequency during normal periods, the effect of target probiotics, and the effect of capsules, respectively. Here, the effect of the capsule indicates the effects of ingesting the capsule, regardless of its content. That is, the effect of the capsule is observed in both the target probiotic and placebo periods. Although $\tilde{\eta}$ is utilized to denote the effect of "the capsule," the proposed model does not rely on the form of the target probiotics or the

placebo. The prior distributions of α, η, and $\tilde{\eta}$ are as follows.

$$\alpha \sim \mathrm{WIP},$$
$$\eta \sim \mathrm{WIP},$$
$$\tilde{\eta} \sim \mathrm{WIP},$$

where WIP denotes the weakly informative prior distribution. We use $\mathrm{Cauchy}(0, 10)$ as the weakly informative prior distribution in this study, where $\mathrm{Cauchy}(x_0, \gamma)$ denotes the Cauchy distribution with the location parameter x_0 and scale parameter γ. Let μ and ν be the lag of effect start and end, respectively. μ and ν are shared by the effects of capsules and probiotics. Specifically, the effects of the capsules and probiotics emerged μ days after the subject ingested the capsule and expired ν days after the subject stopped taking the capsule. Here, we assume that μ and ν are up to several days long using the following prior distributions:

$$\mu \sim \mathrm{DiscreteUniform}(0, \mu_{\max}),$$
$$\nu \sim \mathrm{DiscreteUniform}(0, \nu_{\max}),$$

where $\mathrm{DiscreteUniform}(a, b)$ denotes the discrete uniform distribution with minimum value a and maximum value b, and $\mu_{\max}$ and $\nu_{\max}$ are the maximum values of μ and ν, respectively. We used $\mu_{\max} = \nu_{\max} = 5$ in this study. We modified the segment in which the subject ingested the capsule using μ and ν as follows:

$$d'_1 = d_1 + \mu,$$
$$d'_2 = d_2 + \nu,$$
$$d'_3 = d_3 + \mu,$$
$$d'_4 = d_4 + \nu,$$

where d_1, d_2, d_3, and d_4 represent the start day index of the first capsule, the end day index of the first capsule, the start day index of the second capsule, and the end day index of the second capsule, respectively; and d'_1, d'_2, d'_3, and d'_4 indicate the effect start day index of the first capsule, the effect end day index of the first capsule, the effect start day index of the second capsule, and the effect end day index of the second capsule, respectively. Let O, P, and T be the sets of day indices in the normal, placebo, and target probiotic periods, respectively. O, P, and T are given by

$$O = \{i \mid 1 \leq i < d'_1 \vee d'_2 \leq i < d'_3 \vee d'_4 \leq i\},$$
$$P = \begin{cases} \{i \mid d'_1 \leq i < d'_2\} & C = 0, \\ \{i \mid d'_3 \leq i < d'_4\} & C = 1, \end{cases} \tag{1}$$
$$T = \begin{cases} \{i \mid d'_3 \leq i < d'_4\} & C = 0, \\ \{i \mid d'_1 \leq i < d'_2\} & C = 1, \end{cases} \tag{2}$$

where C indicates the cross-over type, which shows the order of the target probiotic and placebo capsules. The subject was first administered the placebo and then the target probiotic when $C = 0$, and the subject was first administered the target probiotic and then the placebo when $C = 1$. Let β_i be the rate of increase in the cumulative sum of the defecation frequency on the i-th day. β_i depends on the period in which the i-th day lies, as given below:

$$\beta_i = \begin{cases} \exp(\alpha) & i \in O, \\ \exp(\alpha + \tilde{\eta}) & i \in P, \\ \exp(\alpha + \eta + \tilde{\eta}) & i \in T. \end{cases} \tag{3}$$

The intercept of the i-th segment γ_i is defined as

$$\gamma_i = \begin{cases} 0 & i \in G_1, \\ y_{d'_{t-1}-1} - \beta_i(d'_{t-1} - 1) & i \in G_t \wedge t > 1, \end{cases}$$

where G_t is the set of the day indices in the t-th segment. Here, γ_i is calculated such that the regression line passes through the observation point $\left(d'_{t-1} - 1, y_{d'_{t-1}-1}\right)$. This calculation enables the precise evaluation of the increase in the cumulative sum of the defecation frequencies in each segment. γ_i in the first segment equals zero because the cumulative sum of defecation frequencies is 0 before the first day. Here, we considered the placebo effect of defecation frequencies because placebo effects have been confirmed in previous studies [14, 15]. The distribution of the i-th day cumulative sum of defecation frequencies y_i is as follows:

$$y_i \sim \mathrm{Normal}\left(\beta_i i + \gamma_i, \sigma^2\right),$$

where

$$\sigma^2 \sim \mathrm{WIP}_{>\varepsilon}.$$

Here, $\mathrm{WIP}_{>\varepsilon}$ denotes the truncated weakly informative prior distribution, whose domain of definition is $x > \varepsilon$ with a random variable x. We used 0.1 as ε.

2.3 Parameter estimation

We estimated the posterior distribution of the parameters of the proposed model using the No-U-Turn-Sampler (NUTS) [16], which is a Markov chain Monte Carlo (MCMC) method. Because the NUTS can sample only continuous parameters, we estimated the following posterior distribution marginalized with respect to μ and ν:

$$\sum_{\mu}\sum_{\nu} p(\alpha, \eta, \tilde{\eta}, \mu, \nu, \sigma^2 | y_{1:N}, d_{1:4}),$$

where $\cdot_{1:N}$ denotes the set $\{\cdot_i\}_{i=1}^{N}$. We implemented a parameter estimation algorithm using PyStan (`https://github.com/stan-dev/pystan`). We used five chains of MCMC and then sampled parameters 1000 times randomly for each chain and discarded the first half of the samples as burn-in samples, which were supposed to depend on the initial sample. We used 15 as the maximum tree depth in the NUTS algorithm (called the "max_treedepth" option in the PyStan library). The other hyperparameters were set by default.

2.4 Evaluation by scoring improvement of defecation frequency

We defined the following defecation frequency improvement (DFI) score $\mathrm{DFI}(\mu, \nu)$:

$$\mathrm{DFI}(\mu, \nu) \equiv \ln \frac{\frac{1}{|T|}\sum_{i \in T} x_i}{\frac{1}{|P|}\sum_{i \in P} x_i},$$

where x_i is the defecation frequency of the i-th day, and T and P are defined by Eq. (1) and Eq. (2), respectively, for the score parameters μ and ν. The DFI score indicates the log ratio of the defecation frequency in the target probiotic period to that in the placebo period.

2.5 Synthetic data experiment

We generated synthetic datasets and estimated the parameters using these datasets to evaluate the performance of the proposed model. We generated α, η, $\tilde{\eta}$, μ, and ν using the following distribution:

$$\begin{aligned}
\alpha &\sim \mathrm{Normal}(0, 0.1),\\
\eta &\sim \mathrm{Normal}(0, 0.2),\\
\tilde{\eta} &\sim \mathrm{Normal}(0, 0.2),\\
\mu &\sim \mathrm{DiscreteUniform}(0, 5),\\
\nu &\sim \mathrm{DiscreteUniform}(0, 5).
\end{aligned}$$

After computing β_i using α, η, $\tilde{\eta}$, μ, and ν (*cf.* Eq. (3)), the number of days between the l-th and $l+1$-th defecation events for subject $v_l \in \mathbb{R}$ was obtained as follows:

$$v_l \sim \mathrm{Gamma}\left(\frac{1}{{\sigma^{(v)}}^2 \beta_i^2}, {\sigma^{(v)}}^2 \beta_i\right),$$

where $\mathrm{Gamma}(a, b)$ denotes the gamma distribution with the shape parameter a and scale parameter b and ${\sigma^{(v)}}^2$ is the variance of the interval. The mean and variance of this gamma distribution were $1/\beta_i$ and ${\sigma^{(v)}}^2$, respectively. We generated v_l until $\sum_l v_l$ exceeded the number of days in each segment and obtained y_i by transforming the defecation intervals. We randomly generated datasets 1000 times and estimated the posterior distributions of the parameters once for each dataset. We used $(d_1, d_2, d_3, d_4) = (29, 43, 71, 85), (51, 76, 126, 151), (101, 151, 251, 301)$, $N = 85, 151, 301$, ${\sigma^{(v)}}^2 = 0.001, 0.01, 0.1$, and $C = 0$ in one half of the subjects and $C = 1$ in the other half of the subjects for each dataset.

2.6 Real data experiment

We used a real dataset from a previous study [10], which conducted a randomized double-blind controlled cross-over trial. Twenty subjects received *Bifidobacterium longum* capsules in the experiment. Eleven subjects were first administered placebo capsules ($C = 0$), and the remaining subjects were first administered the target probiotic capsules ($C = 1$). $(d_1, d_2, d_3, d_4) = (29, 43, 71, 85)$ was used in this dataset.

2.7 Bayesian beta regression of responder probability on the microbial relative abundances

We conducted regression analysis using the beta regression model [17]. Let r_i be the response probability of the i-th subject. The Bayesian beta regression model represents r_i using the standardized relative abundances of the bacteria in the i-th subject shortly before the start of capsule administration, which is denoted by $\mathbf{m}_i$. $\mathbf{m}_i$ is the D-dimensional vector, and D is the number of the different bacteria. The generative process is as follows:

$$
\begin{aligned}
&\phi \sim \mathrm{WIP}_{>0}, \\
&\lambda \sim \mathrm{WIP}_{>0}, \\
&b_j \sim \mathrm{Normal}(0, \lambda), \\
&\theta_i = \mathrm{logit}^{-1}\left(\mathbf{b}^{\mathrm{T}}\mathbf{m}_i\right), \\
&r_i \sim \mathrm{Beta}(\theta_i\phi, (1-\theta_i)\phi),
\end{aligned}
$$

where ϕ is the precision parameter obtained by reparameterizing the beta distribution parameters. λ is the regularization parameter. $\mathbf{b} = (b_1, \ldots, b_D)^{\mathrm{T}}$ is the regression parameter vector; $\mathrm{logit}^{-1}(\cdot)$ is the inverse-logit function; and $\mathrm{Beta}(a, b)$ denotes the beta distribution with the shape parameters a and b. Because the domain of the definition of the beta distribution does not include zero or one, we added $10^{-5}/-10^{-5}$ to r_i when r_i is zero or one. The same method as in Section 2.3 was used for the parameter estimation.

2.8 Microbiome data

The 16S rRNA sequence data were obtained from the DDBJ DRA(DRA006874). QIIME2 (version 2019.10) was used for the 16S rRNA gene analysis [18]. In the analytical pipeline, sequence data were processed using the DADA2 pipeline for quality filtering and denoising (options: –p-trunc-len-f 150 –p-trunc-len-r 190–p-max-ee-f 3.0 –p-max-ee-r 3.0) [19]. The filtered output sequences were assigned to different taxa using the "qiime feature-classifier classify-sklearn" command with the default parameters. Silva SSU Ref Nr 99 (version 132) was used as the reference database for taxonomy assignment [20]. For the regression analysis, we used only those taxa with a non-zero abundance in at least 15 subjects .

3 Results

3.1 Performance evaluation with synthetic datasets

We evaluated the performance of the proposed model under various conditions using synthetic datasets (*cf.* Section 2.5). To verify the accuracy of η, we compared the estimated and true values (Supplementary Figure S1). In the case of ${\sigma^{(v)}}^2 \leq 0.01$, the proposed model could accurately estimate η. ${\sigma^{(v)}}^2 \leq 0.01$, the standard deviation $\sigma^{(v)} \leq 0.1$, means that defecation events with the one-sigma error are within ± 2.4 hours (*cf.* Section 2.5).

We also verified the estimation accuracy of μ and ν. Figure 2 shows the sum of the probabilities for each true μ and ν to evaluate the uncertainty of the estimation. The diagonal elements in Figure 2def, which shows the results of ν, are high in the ${\sigma^{(v)}}^2 \leq 0.01$ experiments. However, as shown in Figure 2abc, which shows the results for μ, the proposed model tends to overestimate the μ value.

To evaluate the performance improvement by considering the lag between intake and effect periods, we identified responders based on the posterior distributions. Here, we defined responders as the subjects with $\eta > 0$. Figure 3 shows the receiver operating characteristic (ROC) curve of the proposed and base models. We used the proposed model with $\mu_{\max} = \nu_{\max} = 0$ (*cf.* Section 2.2), which does not consider the lag, as the base model. The proposed model outperformed the base model, and the effectiveness of considering the lag was demonstrated in the case where a lag exists. Figure 3 also shows the performance for each threshold of the posterior probability of $\eta > 0$. In the case of ${\sigma^{(v)}}^2 \leq 0.01$, identification with a threshold of 0.95 showed a low false positive rate.

3.2 Responder evaluation using a real dataset

We conducted an experiment using a real dataset (*cf.* Section 2.6). To evaluate the effect of the target probiotic on each subject, we visualized the estimated posterior distribution of η (Figure 4). Subjects MO04, MO05, MO10, and MO16 exhibited high values of η, whereas subjects MO06 and MO18 exhibited low values of η. The 95% Bayesian credible intervals of subjects MO02, MO04, MO05, MO06, MO08, MO10, MO12, MO13, MO16, MO18, and MO23 did not include zero. The posterior distributions of μ and ν are shown in Supplementary Figure S2. We can see that the estimated values of μ and ν vary from person to person.

We also examined the estimated probability that each subject was a responder (Figure 5). We counted the number of samples that satisfied $\eta > 0$ and computed the ratio of the count to the number of all samples as the posterior probability. The probabilities of subjects MO02, MO04, MO05, MO08, MO09, MO10, and MO16 exceeded 0.95. In a previous study, subjects

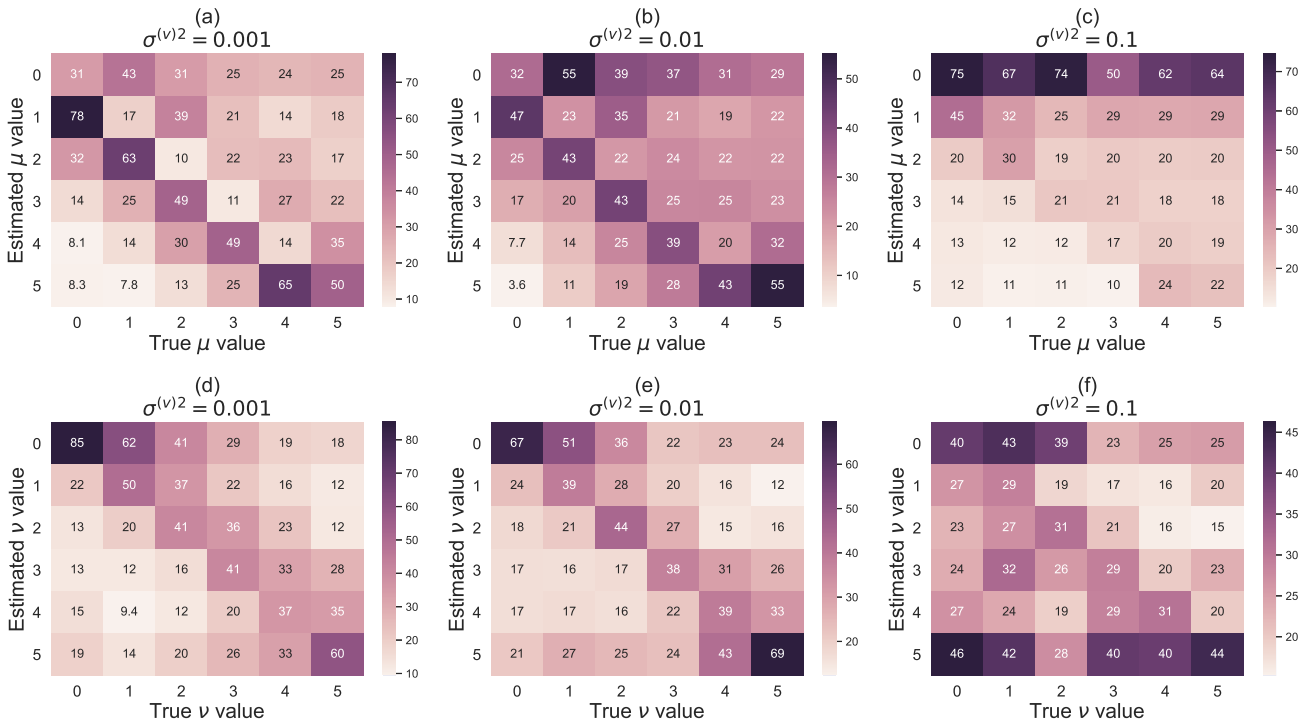

Figure 2: The heat map of the true μ, ν and μ, ν estimated by the proposed model for each synthetic dataset, where the number of observation points is the same as that of the real dataset. The x- and y-axes indicate the true μ, ν and estimated μ, ν values, respectively. Each column represents the sum of the probabilities of each estimate for the subject, whose true value is that in the column. a, b, and c indicate the μ results when ${\sigma^{(v)}}^2 = 0.1$, ${\sigma^{(v)}}^2 = 0.01$, and ${\sigma^{(v)}}^2 = 0.001$, respectively. d, e, and f indicate the ν results when ${\sigma^{(v)}}^2 = 0.1$, ${\sigma^{(v)}}^2 = 0.01$, and ${\sigma^{(v)}}^2 = 0.001$, respectively.

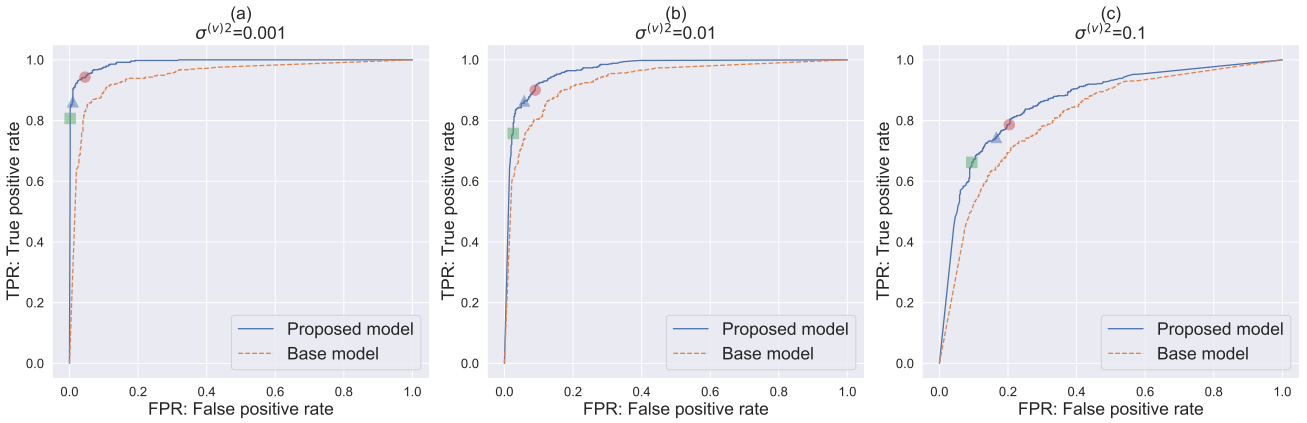

Figure 3: The ROC curve for identifying responders based on the estimated posterior distributions of η in the synthetic datasets of $N = 85$ and $(d_1, d_2, d_3, d_4) = (29, 43, 71, 85)$ using the proposed model ($\mu_{\max} = \nu_{\max} = 5$) and the base model ($\mu_{\max} = \nu_{\max} = 0$). The x- and y-axes indicate the false positive and true positive rates, respectively. The blue and orange dashed lines indicate the results for $\mu_{\max} = \nu_{\max} = 5$ and $\mu_{\max} = \nu_{\max} = 0$, respectively. The red circle, blue triangle, and green square indicate the performance of responder identification with the threshold of the probability of $\eta > 0$ 0.5, 0.7, and 0.95, respectively, when $\mu_{\max} = \nu_{\max} = 5$.

MO04, MO05, and MO10 were reported as responders, whereas subject MO16 was a non-responder [10]. Supplementary Figure S3 shows the posterior distributions of η estimated by the base model, which did not consider the lag ($\mu_{\max} = \nu_{\max} = 0$). The Bayesian credible interval of subject MO16 also did not include zero, but the median was estimated to be lower than that of the proposed model. The results for subjects MO22 and MO24 showed large differences between the proposed and base models. The median values of the base model were larger than those of the proposed model, and their identification of responders based on 95% Bayesian credible intervals led to different conclusions (Figure 5 and Supplementary Figure S4).

To verify the consistency between the posterior distributions and the used dataset, we evaluated the improvement in the defecation frequency using scoring (*cf.* Section 2.4). Figure 6 shows the DFI score of subjects where $\mu = 0 \ldots 5$ and $\nu = 0 \ldots 5$.

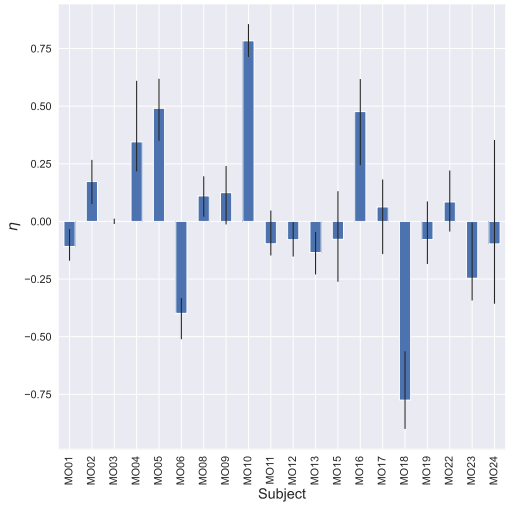

Figure 4: Estimated posterior distributions of η for each subject. The x- and y-axes indicate the subjects and η values, respectively. The bar shows the median of the posterior distribution. The error bars represent the 2.5% and 97.5% percentiles.

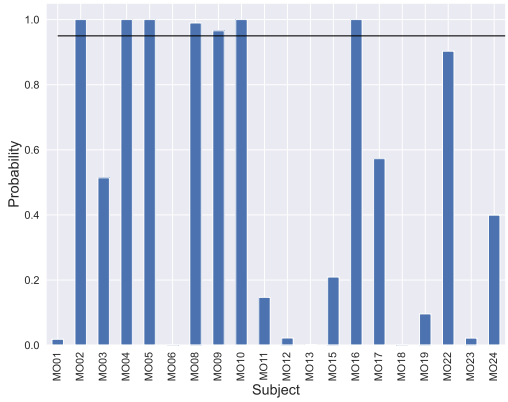

Figure 5: The probability that each subject is a responder based on the posterior distribution. The x- and y-axes indicate the subjects and probabilities of $\eta > 0$, respectively. The horizontal line indicates that the probability equals 0.95.

While the DFI scores of $\mu = 0$ and $\nu = 0$ for subject MO24, which was identified as a responder by the base model, were 0.18, the DFI scores for $\mu \neq 0$ and $\nu \neq 0$ were less than 0. Therefore, subject MO24 did not show an improvement in the defecation frequency when a lag in the effect period existed, and the proposed model reflected the specifications of subject MO24. We also examined the fit of the predictive distribution to the data set (Figure 7). We observed that using the cumulative sum enabled the consideration of the uncertainty caused by uneven defecation frequencies. For example, the defecation frequencies of MO04 and MO12 were comparable, but the uncertainty was estimated to be larger for MO04 because of the uneven defecation frequencies.

To investigate the relationship between the response to probiotics and gut microbiota, we performed Bayesian beta regression of the responder probability on the microbial abundance features before the target probiotic periods. Figure 8 shows the posterior distribution of the regression parameters. As in the previous study [10], the negative effect of *Agathobacter* was estimated. The 95% Bayesian credible intervals for all regression parameters included zero.

4 Discussion

In this study, we proposed a model for estimating the improvement in defecation frequencies using cross-over trial datasets and considering the lag between intake and effect periods. Using synthetic datasets, we verified that the proposed model could identify responders better than the base model. In the real dataset experiments, we identified seven responders based on the probability of $\eta > 0$. The base model, which assumed that the lag did not exist, identified subjects MO22 and MO24 as responders in addition to those identified by the proposed model. Subjects MO22 and MO24 did not have high DFI scores when $\mu \neq 0$ and $\nu \neq 0$ (Figure 6). The proposed model reflected these observations. The proposed model was suggested to eliminate uncertain responders regarding the lag between intake and effect periods. In the regression analysis of the responder probabilities on the microbial relative abundances before target probiotic intake, we found that *Agathobacter* had a negative

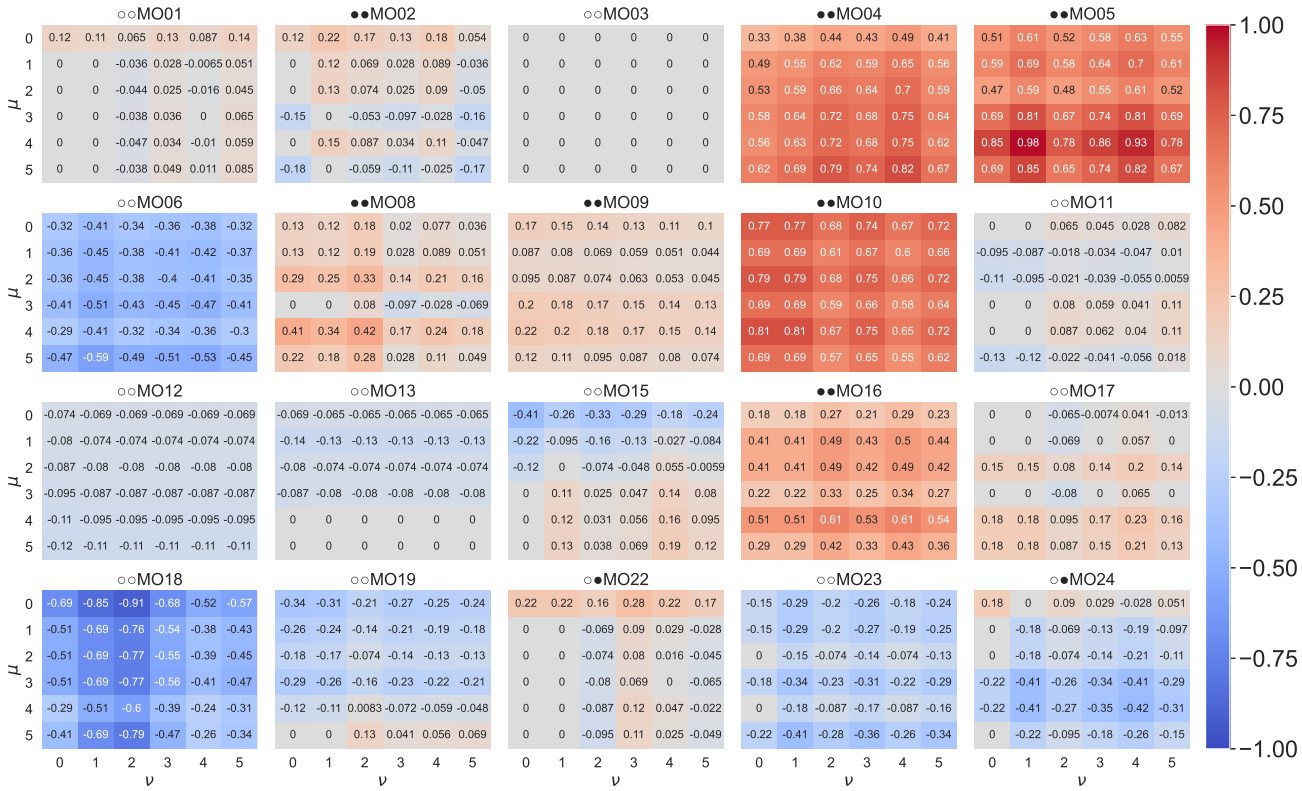

Figure 6: DFI scores (*cf.* Section 2.4) of all subjects. The title of each panel indicates the subject and the result of the responder identification by the proposed and base models. The left and right circles indicate the proposed and base model results, respectively. The filled circle indicates that the subject is identified as a responder, and the open circle indicates that the subject is not. The x- and y-axes indicate ν and μ of the score parameters, respectively. Each value indicates the score. A darker color indicates a higher score.

effect. This result for *Agathobacter* is consistent with that of the previous study [10]. However, we could not conclude any microbial effect based on the 95% Bayesian credible intervals.

One of the benefits of using statistical Bayesian models is its adaptability and the potential for extensive customization. If our model is desired to be applied to other phenotype datasets, it can be employed with a slightly modified structure. For example, consider a blood pressure value case. y_i indicates the blood pressure value, which is continuous and not cumulative. In this case, the y_i distribution of the form

$$y_i \sim \mathrm{Normal}(\beta_i, \sigma^2),$$

can represent the continuous values for days in each segment. Therefore, although the proposed model in this study was applied to the cumulative sum of defecation frequencies, the key idea of the model is supposed to be widely applicable in cross-over trial datasets.

We used the same lag of the start/end day for the placebo and the target probiotic periods. However, these two types of lags may differ because of their sources. While the lag in the target probiotic period is likely to be caused by physical factors (digestion and changes in physical conditions), the lag in the placebo period is likely to be caused by cognitive factors [21]. Therefore, introducing different lag parameters for the placebo and target probiotic periods may enable better estimation. However, adding these parameters can render the estimation computationally expensive.

We used uniform distributions with fixed hyperparameters as prior distributions of μ and ν. There are several options for the prior distribution. Setting prior distributions based on literature enables a more accurate estimation of parameters. In addition, the covariance between μ and ν can reflect the consistency of μ and ν in a subject if a covariate between μ and ν can be assumed.

μ and ν play key roles in the proposed model. In the synthetic data experiments (Section 3.1), the estimation performance of μ and ν was not very accurate. However, the proposed model is effective for estimating responders, as seen in the synthetic data experiments (Figure 3). This is because considering all cases of (μ, ν) contributes to the detection of responders when there is a lag in the effect period. However, μ and ν may not always be indispensable. The effects of μ and ν are limited for the long-term datasets because the number of lag days is small relative to the number of days in the trial. Indeed, the difference between the base model and the proposed model is smaller for synthetic long-term datasets containing observations made under similar conditions (Supplementary Figure S5 and Supplementary Figure S6). Nevertheless, consideration of lag is still significant due to the following reasons: first, the proposed model still outperforms the base model in long-term datasets

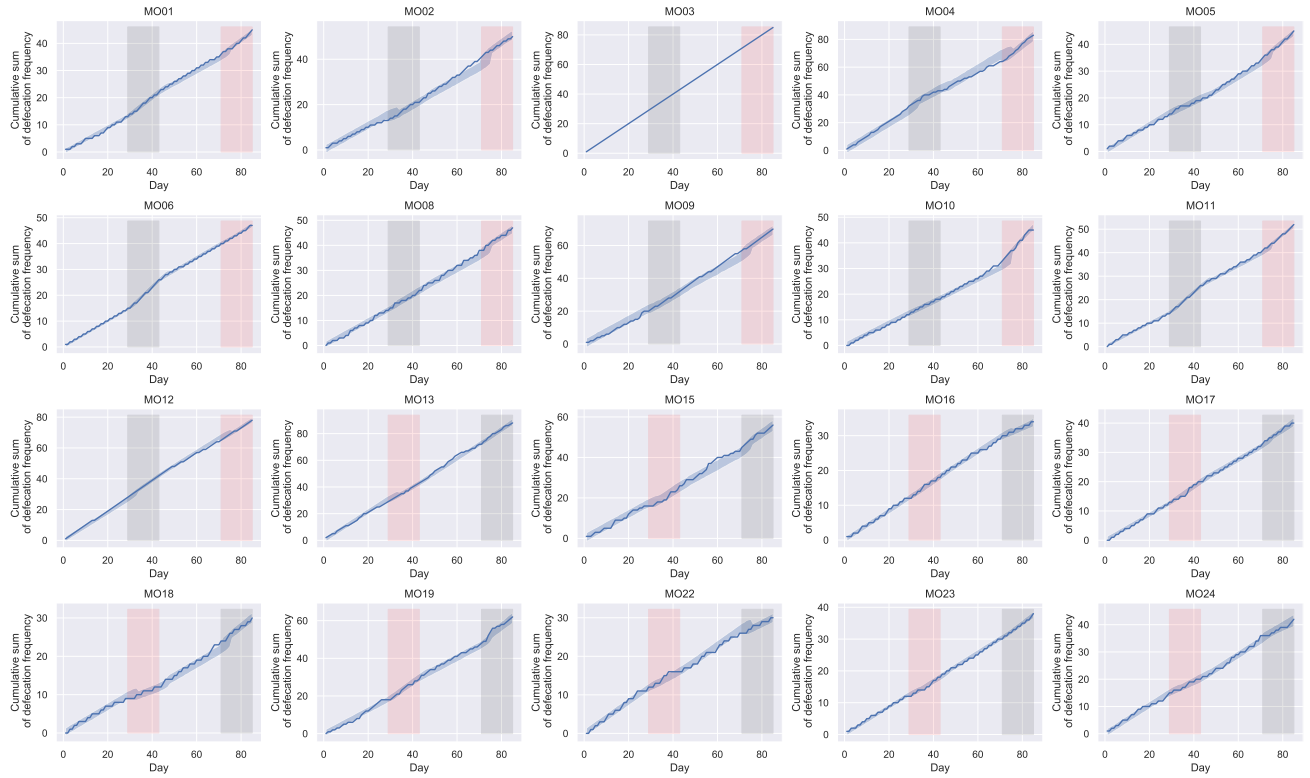

Figure 7: Data used for the analysis and predictive distributions. Each plot shows each subject. The x- and y-axes indicate the day and cumulative sum of the defecation frequencies, respectively. The blue line and the blue area indicate real data and the 2.5% and 97.5% percentiles of predictive distributions, respectively. The red and black areas indicate the target probiotic and placebo intake periods, respectively.

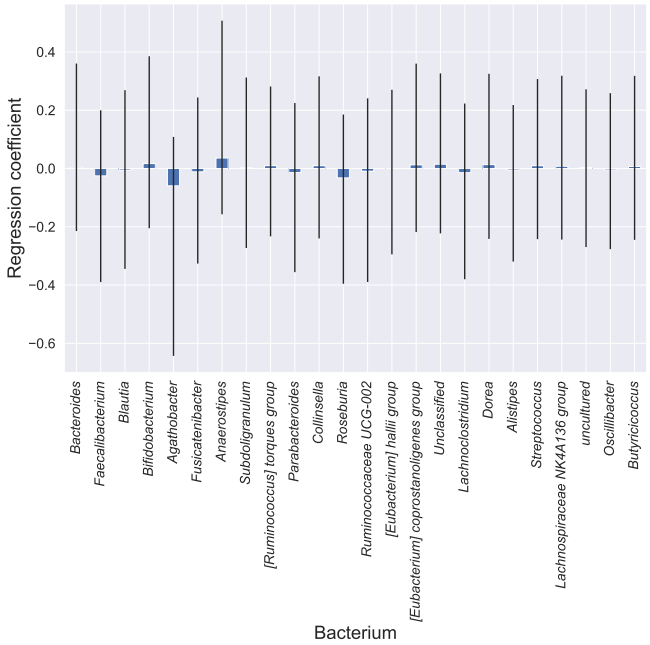

Figure 8: Posterior distributions of the coefficients of Bayesian beta regression. The x- and y-axes indicate bacteria and regression coefficients, respectively. The bar shows the median of the posterior distribution. The error bars represent the 2.5% and 97.5% percentiles.

(Supplementary Figure S5 and Supplementary Figure S6), and second, experiments are frequently limited to a brief duration for economic reasons.

There is a limitation to determining responders based only on defecation frequencies, which is suggested to be unreliable by the U.S. Department of Health and Human Services Food and Drug Administration [22]. According to them, the identification

of responders needs to be evaluated based on defecation frequency and abdominal pain intensity. Therefore, deterministic estimation of responders may lead to wrong conclusions. We believe that the responder estimation based on the posterior distribution of η conducted in this study enables us to consider the uncertainty of the estimation and contribute to solving this problem.

In conclusion, the proposed model effectively addresses the lag between intake and effect periods. The model demonstrated strong performance when lag was present.

Author contributions

Shion Hosoda: Conceptualization, Methodology, Software, Investigation, Validation, Visualization, Writing-Original Draft. **Yuichiro Nishimoto:** Software, Investigation, Writing - Review & Draft. **Yohsuke Yamauchi:** Software, Investigation, Writing - Review & Draft. **Takuji Yamada:** Investigation, Writing - Review & Draft. **Michiaki Hamada:** Investigation, Supervision, Writing - Review & Draft.

Acknowledgments

Computations were partially performed with the NIG supercomputer at ROIS National Institute of Genetics and the supercomputer at Human Genome Center, the Institute of Medical Science, the University of Tokyo. We would like to thank Editage (www.editage.jp) for English language editing.

Data availability

Stan and Python source codes are available at `https://github.com/shion-h/LagBasedResponderIdentifier`.

Funding

This work was supported by JSPS/MEXT KAKENHI (Grant Number JP19J20117 to SH, JP20H00624, JP19H01152, and JP18KT0016 to MH).

Supplementary data

Supplementary data, including Figures S1, S2, S3, S4, and S5, are available.

Declarations

The randomized controlled trial whose dataset was used in this study was conducted with the approval of the clinical trial ethics review committee of Chiyoda Paramedical Care Clinic.

References

[1] Anna Chmielewska and Hania Szajewska. Systematic review of randomised controlled trials: Probiotics for functional constipation. *World Journal of Gastroenterology : WJG*, 16(1):69–75, January 2010.

[2] Katarzyna Wojtyniak and Hania Szajewska. Systematic review: Probiotics for functional constipation in children. *European Journal of Pediatrics*, 176(9):1155–1162, September 2017.

[3] Eirini Dimidi, Stephanos Christodoulides, Konstantinos C. Fragkos, S. Mark Scott, and Kevin Whelan. The effect of probiotics on functional constipation in adults: A systematic review and meta-analysis of randomized controlled trials. *The American Journal of Clinical Nutrition*, 100(4):1075–1084, October 2014.

[4] Petia Kovatcheva-Datchary, Anne Nilsson, Rozita Akrami, Ying Shiuan Lee, Filipe De Vadder, Tulika Arora, Anna Hallen, Eric Martens, Inger Björck, and Fredrik Bäckhed. Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella. *Cell Metabolism*, 22(6):971–982, December 2015.

[5] Tal Korem, David Zeevi, Niv Zmora, Omer Weissbrod, Noam Bar, Maya Lotan-Pompan, Tali Avnit-Sagi, Noa Kosower, Gal Malka, Michal Rein, Jotham Suez, Ben Z. Goldberg, Adina Weinberger, Avraham A. Levy, Eran Elinav, and Eran Segal. Bread Affects Clinical Parameters and Induces Gut Microbiome-Associated Personal Glycemic Responses. *Cell Metabolism*, 25(6):1243–1253.e5, June 2017.

[6] Nicole J. Kellow, Melinda T. Coughlan, and Christopher M. Reid. Metabolic benefits of dietary prebiotics in human subjects: A systematic review of randomised controlled trials. *British Journal of Nutrition*, 111(7):1147–1161, April 2014.

[7] Farnaz Abdollahi, Emily D. Case Lazarro, Molly Listenberger, Robert V. Kenyon, Mark Kovic, Ross A. Bogey, Donald Hedeker, Borko D. Jovanovic, and James L. Patton. Error Augmentation Enhancing Arm Recovery in Individuals With Chronic Stroke: A Randomized Crossover Design. *Neurorehabilitation and Neural Repair*, 28(2):120–128, February 2014.

[8] Brian Budgell and Barbara Polus. The Effects of Thoracic Manipulation on Heart Rate Variability: A Controlled Crossover Trial. *Journal of Manipulative and Physiological Therapeutics*, 29(8):603–610, October 2006.

[9] Chi-Yeon Lim and Junyong In. Considerations for crossover design in clinical study. *Korean Journal of Anesthesiology*, 74(4):293–299, 2021.

[10] Yuya Nakamura, Shinya Suzuki, Shinnosuke Murakami, Yuichiro Nishimoto, Koichi Higashi, Naoki Watarai, Junpei Umetsu, Chiharu Ishii, Yutaro Ito, Yuka Mori, et al. Integrated gut microbiome and metabolome analyses identified fecal biomarkers for bowel movement regulation by bifidobacterium longum bb536 supplementation: A rct. *Computational and structural biotechnology journal*, 20:5847–5858, 2022.

[11] Francesco Asnicar, Emily R. Leeming, Eirini Dimidi, Mohsen Mazidi, Paul W. Franks, Haya Al Khatib, Ana M. Valdes, Richard Davies, Elco Bakker, Lucy Francis, Andrew Chan, Rachel Gibson, George Hadjigeorgiou, Jonathan Wolf, Timothy D. Spector, Nicola Segata, and Sarah E. Berry. Blue poo: Impact of gut transit time on the gut microbiome using a novel marker. *Gut*, 70(9):1665–1674, September 2021.

[12] Niall G. Vine, Winston D. Leukes, and Horst. Kaiser. In vitro growth characteristics of five candidate aquaculture probiotics and two fish pathogens grown in fish intestinal mucus. *FEMS Microbiology Letters*, 231(1):145–152, February 2004.

[13] Abigail J. Johnson, Pajau Vangay, Gabriel A. Al-Ghalith, Benjamin M. Hillmann, Tonya L. Ward, Robin R. Shields-Cutler, Austin D. Kim, Anna Konstantinovna Shmagel, Arzang N. Syed, Jens Walter, Ravi Menon, Katie Koecher, and Dan Knights. Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans. *Cell Host & Microbe*, 25(6):789–802.e5, June 2019.

[14] Slavenka Kam-Hansen, Moshe Jakubowski, John M Kelley, Irving Kirsch, David C Hoaglin, Ted J Kaptchuk, and Rami Burstein. Labeling of medication and placebo alters the outcome of episodic migraine attacks. *Science translational medicine*, 6(218):218ra5, 2014.

[15] Sigurlaug H Hafliðadóttir, Carsten B Juhl, Sabrina M Nielsen, Marius Henriksen, Ian A Harris, Henning Bliddal, and Robin Christensen. Placebo response and effect in randomized clinical trials: meta-research with focus on contextual effects. *Trials*, 22(1):1–15, 2021.

[16] Matthew D. Hoffman and Andrew Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(47):1593–1623, 2014.

[17] Silvia Ferrari and Francisco Cribari-Neto. Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, 31(7):799–815, August 2004.

[18] Evan Bolyen, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian C. Abnet, Gabriel A. Al-Ghalith, Harriet Alexander, Eric J. Alm, Manimozhiyan Arumugam, Francesco Asnicar, Yang Bai, Jordan E. Bisanz, Kyle Bittinger, Asker Brejnrod, Colin J. Brislawn, C. Titus Brown, Benjamin J. Callahan, Andrés Mauricio Caraballo-Rodríguez, John Chase, Emily K. Cope, Ricardo Da Silva, Christian Diener, Pieter C. Dorrestein, Gavin M. Douglas, Daniel M. Durall, Claire Duvallet, Christian F. Edwardson, Madeleine Ernst, Mehrbod Estaki, Jennifer Fouquier, Julia M. Gauglitz, Sean M. Gibbons, Deanna L. Gibson, Antonio Gonzalez, Kestrel Gorlick, Jiarong Guo, Benjamin Hillmann, Susan Holmes, Hannes Holste, Curtis Huttenhower, Gavin A. Huttley, Stefan Janssen, Alan K. Jarmusch, Lingjing Jiang, Benjamin D. Kaehler, Kyo Bin Kang, Christopher R. Keefe, Paul Keim, Scott T. Kelley, Dan Knights, Irina Koester, Tomasz Kosciolek, Jorden Kreps, Morgan G. I. Langille, Joslynn Lee, Ruth Ley, Yong-Xin Liu, Erikka Loftfield, Catherine Lozupone, Massoud Maher, Clarisse Marotz, Bryan D. Martin, Daniel McDonald, Lauren J. McIver, Alexey V. Melnik, Jessica L. Metcalf, Sydney C. Morgan, Jamie T. Morton, Ahmad Turan Naimey, Jose A. Navas-Molina, Louis Felix Nothias, Stephanie B. Orchanian, Talima Pearson, Samuel L. Peoples, Daniel Petras, Mary Lai Preuss, Elmar Pruesse, Lasse Buur Rasmussen, Adam Rivers, Michael S. Robeson, Patrick Rosenthal, Nicola Segata, Michael Shaffer, Arron Shiffer, Rashmi Sinha, Se Jin Song, John R. Spear, Austin D. Swafford, Luke R. Thompson, Pedro J. Torres, Pauline Trinh, Anupriya Tripathi, Peter J. Turnbaugh, Sabah Ul-Hasan, Justin J. J. van der Hooft, Fernando Vargas, Yoshiki Vázquez-Baeza, Emily Vogtmann, Max von Hippel, William Walters, Yunhu Wan, Mingxun Wang, Jonathan Warren, Kyle C. Weber, Charles H. D. Williamson, Amy D. Willis, Zhenjiang Zech Xu, Jesse R. Zaneveld, Yilong Zhang, Qiyun Zhu, Rob Knight, and J. Gregory Caporaso. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8):852–857, August 2019.

[19] Benjamin J. Callahan, Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7):581–583, July 2016.

[20] Christian Quast, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1):D590–D596, January 2013.

[21] Donald D. Price, Damien G. Finniss, and Fabrizio Benedetti. A Comprehensive Review of the Placebo Effect: Recent Advances and Current Thought. *Annual Review of Psychology*, 59(1):565–590, 2008.

[22] US Food, Drug Administration, et al. Guidance for industry irritable bowel syndrome–clinical evaluation of products for treatment, 2010.