**SDC 5: Details of Statistical Models**

*Notation*

Let be the outcome and be the treatment taking on values of , where the three treatments correspond to the baseline diet, SCD and MSCD. Let index the participants, and index the days in the trial on which repeated measurements on a single participant are taken. Let be the set of treatments that participant takes (some individuals who withdrew before completing the first crossover will only have data on one or two diets).

*Individual N-of-1 Trials*

The IBD Symptoms and PROMIS® Pain Interference outcomes are modelled for an individual trial as

where , are the diet-specific intercepts; is the indicator function equal to one when and zero when are autocorrelated random deviations from each diet mean with the lag-1 autocorrelation; are independent, normally distributed random errors with mean zero and variance on day 1 for stationarity and from day 2.

We assume that daily stool frequency and consistency follow Poisson and Bernoulli distributions, respectively, where we do not adjust for autocorrelation, and the models for individual trials are

where is the link function. We choose non-informative priors that are uniformly distributed on the possible values of the model parameters and link functions:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Outcome |  |  |  |  |
| IBD Symptoms (Parent) | Uniform | Uniform | Uniform | - |
| IBD Symptoms (Child) | Uniform |
| Pain Interference (Parent) | Uniform |
| Pain Interference (Child) | Uniform |
| Stool Frequency | N | - | - | log |
| Stool Consistency | N | - | - | logit |

A mean difference of 3 points for IBD Symptoms and PROMIS® Pain Interference and a relative change of 10% and 20% for stool frequency and consistency, respectively, are considered clinically meaningful. Therefore, for comparison between diet and , we report the posterior probability of the following quantities falling into the corresponding intervals and their posterior median and 95% credible interval:

|  |  |  |
| --- | --- | --- |
| Outcome | Effect estimate | Intervals |
| IBD Symptoms |  |  |
| Pain Interference |
| Stool Frequency |  |  |
| Stool Consistency |  |  |

*Aggregate Analysis*

We fit models to estimate the average treatment effects in full completers, early completers and withdrawals separately, and use the modeling results in each group to impute the missing measurements assuming them missing at random. We then estimate the average treatment effects across all individuals using the imputed datasets. Below we detail the models we fit in each group, the imputation process and the models we fit across all individuals.

For IBD Symptoms and PROMIS® Pain Interference in each group, we fit the contrast-based model describing observed arms adjusting for autocorrelation,1

where is the fixed participant-specific intercept; is the random diet effect comparing diet to the usual diet for participant (usual diet, which is available for all participants, is taken as the reference diet and labelled k=1); is the average effect comparing diet to diet 1; is the random component for the random diet effect; is the variance for the random diet effects. We assume a common variance for random diet effects comparing any two of the three diets, which implies a compound symmetry variance structure with the correlation among random effects being 0.5. We define . Other model components are as for the individual-participant models.

For daily measured stool frequency and consistency in each group, we fit the same contrast-based model with random diet effects except that we do not adjust for autocorrelation as in the individual-participant model,

For fecal calprotectin in each group, we model the outcome on the log scale because the distribution on the original scale is left-skewed. Additionally, since fecal calprotectin was collected once in each diet period, we fit an arm-based model with random diet-specific intercepts,1

corr

corr

where is the random diet-specific intercept for participant ; is the average response on diet ; is the random component to the random diet-specific intercept; is the variance for the random intercepts; is the correlation between random intercepts. The limited number of fecal calprotectin measurements per participant made it hard to distinguish between intraclass correlation in the mixed models and autocorrelation among repeated measurements so we assumed the to be independent. Moreover, because they were spaced 8 weeks apart, they were less likely to be correlated. Because the number of participants in each group is small, we do not adjust for stratification factors in the models for each group.

We impute missing measurements based on the modeling results in each group to obtain the datasets for estimating the average treatment effects across all individuals. We assume that the measurements in each group are missing at random. Both intermittent missing measurements and missing measurements due to dropout are imputed to ensure that all participants have complete data throughout the study period. For weekly measured IBD Symptoms and PROMIS® Pain Interference and daily measured stool frequency and consistency, we impute the missing measurements so that all participants have complete measurements for at least one week on baseline diet and for at least six weeks in each SCD or MSCD period (i.e., each participant will have complete data for at least 25 weeks). For fecal calprotectin, we impute the missing measurements so that all participants have one measurement in baseline and each diet period (i.e., each participant will have five measurements). Five datasets are imputed for each outcome for multiple imputation.

For the models across all individuals, the contrast-based models for IBD symptoms, pain interference, stool frequency and consistency implicitly adjust for stratification factors used in randomization (i.e., clinical site and disease condition) by including a fixed intercept for each patient; we adjust for disease type in the random-intercept model for fecal calprotectin. Because of the sparseness of the clinical sites, we do not adjust for clinical sites in the model for fecal calprotectin. Specifically, we included an additional term for disease type in the second-level model for fecal calprotectin across all individuals,

We choose the following priors for the parameters in the models:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Outcome |  |  |  |  |
| IBD Symptoms | Uniform | Uniform | Uniform | Uniform |
| Pain Interference | Uniform | Uniform |
| Fecal Calprotectin | N | - |
| Stool Frequency | N | N | - |
| Stool Consistency | N | N | Uniform |
| Fecal Calprotectin | N | - | Uniform | Uniform |

Additionally, we choose the following link functions and outcome distributions for stool frequency and consistency:

|  |  |  |
| --- | --- | --- |
| Outcome |  | Outcome Distribution |
| Stool Frequency | log | Poisson |
| Stool Consistency | logit | Bernoulli |

For comparison between diet and , we report the posterior probability of the following quantities falling into the corresponding intervals and their posterior median and 95% credible interval:

|  |  |  |
| --- | --- | --- |
| Outcome | Effect estimate | Intervals |
| IBD Symptoms |  |  |
| Pain Interference |
| Stool Frequency |  |  |
| Stool Consistency |  |
| Fecal Calprotectin |  | - |

For fecal calprotectin, we also present the posterior geometric mean for each diet and the corresponding 95% credible interval.

**Sample Size Calculations**

The calculations leading to sample size formulae for the number of individuals in multi-crossover trials with multiple measurements per period needed to detect an average treatment effect (ATE) of a specific size in order to achieve a specific power and significance level are detailed below.

In brief, the power, 1-β, to detect a difference of δ between the responses in the two treatment groups is calculated as

where . Here F(a,b,c) is the distribution function of the non-central t-distribution with n-1 degrees of freedom, significance level α and non-centrality parameter ; is the variance of treatment effects across individuals; is the within-person variance; K is the number of time blocks when patients take a treatment (i.e. number of crossovers including first period); L is the number of measurements per treatment period; ρ is the correlation between consecutive measurements within-person; and n Is the number of individuals.

For the comparison of two IBD diets, we have estimated power based on the PROMIS Pain Interference scale that is measured weekly and is one of the primary outcomes for this aim. The PROMIS pain Interference Scale is standardized by a T-score measure to have mean of 50 and between-patient standard deviation . In a sample of 53 teenage children with IBD who completed the PROMIS Pain Interference scale weekly between 2 and 91 times each (total of 1736 observations), we calculated  (which is close to the assumed 10) and Assuming, the standard deviation of the crossover effect which does not incorporate the number of patients IJ is if the observations are independent and  for AR(1) errors. The study design consists of 4 periods of 8 weeks each, during which time patients will cross-over to each diet twice (K=2). In order to account for carryover, we will introduce a 1-week washout period analytically by disregarding the first week's measurements for analysis. This will leave L=7 measurements per period for a total of 14 weeks data on each treatment. Assuming a two-sided significance level of 0.05, the tables below shows the minimum detectable difference between treatment groups for different and when K=2, L =7, for N = 50, 100 and with Power = 0.8, 0.9.

N = 100 and power = 0.9

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | 0 | 0.2 | 0.5 | 0.6 | 0.7 |
| 0 | 0.74 | 0.91 | 1.29 | **1.48** | 1.77 |
| 1 | 0.81 | 0.97 | 1.33 | **1.52** | 1.80 |
| 5 | 1.80 | 1.87 | 2.08 | **2.21** | 2.41 |

N = 50 and power = 0.9

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | 0 | 0.2 | 0.5 | 0.6 | 0.7 |
| 0 | 1.06 | 1.30 | 1.84 | **2.12** | 2.52 |
| 1 | 1.16 | 1.38 | 1.90 | **2.17** | 2.57 |
| 5 | 2.57 | 2.67 | 2.97 | **3.16** | 3.44 |

N = 100 and power = 0.8

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | 0 | 0.2 | 0.5 | 0.6 | 0.7 |
| 0 | 0.64 | 0.79 | 1.11 | **1.28** | 1.53 |
| 1 | 0.70 | 0.84 | 1.15 | **1.31** | 1.55 |
| 5 | 1.55 | 1.62 | 1.80 | **1.91** | 2.08 |

N = 50 and power = 0.8

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | 0 | 0.2 | 0.5 | 0.6 | 0.7 |
| 0 | 0.92 | 1.12 | 1.59 | **1.83** | 2.18 |
| 1 | 1.00 | 1.19 | 1.64 | **1.88** | 2.22 |
| 5 | 2.22 | 2.31 | 2.57 | **2.73** | 2.97 |

A minimally important difference for PROMIS Pain Interference scale has been reported as 2-3 points (with 3 being reported by parents and adolescents). This corresponds to an effect size of 0.3 given the between-person standard deviation of 10 points. As can be seen from the tables above all the combinations give minimal detectable differences of 3 points or fewer except when and when the minimal difference is above 3 to obtain 90% (but not 80% power) with N = 50. It is highly unlikely that the standard deviation of the treatment effect would be as large as 5 (which would correspond to at least a 20-point range in treatment effects across the range of individuals), so effectively all scenarios would retain power of at least 90% to detect a 3-point change for sample sizes of 50.

**Reference**

1 White, I.R., Turner, R.M., Karahalios, A. and Salanti, G., 2019. A comparison of arm‐based and contrast‐based models for network meta‐analysis. *Statistics in medicine*, *38*(27), pp.5197-5213.