Supplementary File 1

Methods:

Trial registration:
This trial was registered at ClinicalTrials.gov under identifier NCT04132427 on 11 October 2019, entitled *MTT for Children With Both Pitt Hopkins Syndrome and Gastrointestinal Disorders* and link https://clinicaltrials.gov/study/NCT04132427

Participants Recruitment criteria
Inclusion Criteria
1) Children ages 7-17 years with Pitt Hopkins Syndrome (verified by genetic testing).
2) GI disorder as defined below that has lasted for at least 2 years.
3) No changes in medications, supplements, diet, or therapies in last 2 months, and no intention to change them during the Parts 1 and 2 of the clinical trial.
4) Ability to swallow pills (without chewing) - this will be verified by the study staff during the first eligibility visit by administering a placebo capsule and ensuring that the participant can swallow it without chewing.
5) Review of last two years of medical records by the study physician.

Exclusion Criteria
1) Antibiotics in last 3 months (topical antibiotics are allowed).
2) Probiotics in last 2 months, or fecal transplant in last 12 months.
3) Tube feeding.
4) Severe gastrointestinal problems that require immediate treatment (life-threatening).
5) Ulcerative Colitis, Crohn's Disease, diagnosed Celiac Disease, Eosinophilic Gastroenteritis, or similar conditions.
6) Unstable, poor health (based on study physician's opinion).
7) Recent or scheduled surgeries.
8) Current participation in other clinical trials.
9) Females who are pregnant or who are at risk of pregnancy and sexually active without effective birth control. We will conduct a pregnancy test on all female participants as part of the screening and at each clinical visit.
10) Allergy or intolerance to vancomycin or magnesium citrate.
11) Clinically significant abnormalities at baseline on two blood safety tests: Comprehensive Metabolic Panel, and Complete Blood Count with Differential. Note that some abnormalities may occur due to PTHS, so only those likely to significantly increase risk in this study would be grounds for exclusion, at the discretion of the study physician. See detailed discussion at end of this section on Interpreting Laboratory Results. re. Eligibility for Admission to Study.
12) Evidence of significant impairment of immune system, or taking medications that can compromise the immune system, and thus increase risk if expose to multiple-drug resistant bacteria.

The rational for exclusions 1-2 is that they interfere with gut flora.
The rationale for exclusion 3 is that those individuals are probably less likely to respond to the proposed treatment.
The rationale for exclusions 4-7 is that those individuals are at higher risk of safety problems with MTT.
The rational for exclusion 8 is that participation in other trials would interfere with the results of this one.
The rationale for exclusion 9 is to avoid risk to fetuses.
The rationale for exclusion 10 is to avoid known allergic reactions to study medications.

The rationale for exclusions 11 is to ensure that participants are in adequate health (aside from chronic GI and PTHS-related problems) at the start of the study.

Definition of a GI Disorder (for this study):

On the Daily Stool and Symptom Record, five or more abnormal days out of 14 days, where an abnormal day is defined as a day which involves one or more of the following symptoms:

1) Abnormal stool: type 1-2 (hard) or type 6-7 (soft/liquid) on the Bristol Stool Scale
2) Three or more bowel movements in 1 day
3) No bowel movement that day.
4) Significant abdominal/gastrointestinal pain.
5) GI medication or treatment is required (such as laxative, enema, etc.)

Dosage regime

Magnesium citrate bowel preparation was administered orally using a commercially available solution with a concentration of 1.75 g per 30 mL. Dosing was determined according to an age-stratified clinical protocol. Children aged 5–8 years received 4 mL/kg body weight. For older age groups, a fixed total volume was administered as follows: 150 mL for ages 8–12 years, 200 mL for ages 12–15 years, and 300 mL for ages 15–18 years.

1. Initial MTP-101C FM Oral Dose: The dosage for the first 4 four days was approximately 5x 10^{11} cells/day. It was administered as two capsules in the morning and 3 capsules in the evening.

2. Maintenance MTP-101C FM Oral Dose: The maintenance dose for the next 12 weeks was approximately 1x 10^{11} cells (1 capsule) every 4 four days.

Assessments of gastrointestinal and PTHS-related symptoms

Daily Stool Records (DSR) We collected DSR (primary outcome), recorded by participants or parents that rates the stool using the Bristol Stool Form scale (1-2 = very hard, 3-5 normal, 6-7 = liquid) over 14 days (Kang et al., 2019). Days without a bowel movement, or with 3 or more bowel movements, will be counted as an abnormal day. Days requiring a GI medication/treatment, such as a laxative or enema, will also be counted as an abnormal day. Days with significant abdominal pain would also be counted as an abnormal day. The reason for this scale is that most PTHS patients have constipation, but some have diarrhea, and some alternate between the two. This simple form allows us to describe either type of problem as an "abnormal" day, and hence use a single number to describe constipation and/or diarrhea symptoms.

Subscales will include the percentage of days of types 1-2 (constipation), types 6-7 (diarrhea), days with no stool, days with 3 or more stools, days requiring GI medication/treatment, and days with significant abdominal pain.

In Phase 1 study for autism (Kang et al., 2017), we found that the DSR was easy for patients to understand and comply with, and it was able to detect significant changes due to MTT treatment.

Reference for DSR:

Kang, DW., Adams, J.B., Gregory, A.C. et al. Microbiota Transfer Therapy alters gut ecosystem and improves gastrointestinal and autism symptoms: an open-label study. *Microbiome* 5, 10 (2017). https://doi.org/10.1186/s40168-016-0225-7

Kang, DW., Adams, J.B., Coleman, D. et al. Long-term benefit of Microbiota Transfer Therapy on autism symptoms and gut microbiota. *Sci Rep* 9, 5821 (2019). https://doi.org/10.1038/s41598-019-42183-0

Bristol Stool Form Scale

THE BRISTOL STOOL FORM SCALE		
Type 1		Separate hard lumps, like nuts
Type 2		Sausage-like but lumpy
Type 3		Like a sausage but with cracks in the surface
Type 4		Like a sausage or snake, smooth and soft
Type 5		Soft blobs with clear-cut edges
Type 6		Fluffy pieces with ragged edges, a mushy stool
Type 7		Watery, no solid pieces

Clinical Global Impression (CGI) The CGI is a widely used tool for assessing severity of symptoms, and changes after treatment. The CGI- Severity Scale (CGI-S) was used at baseline, and after treatment evaluations included the CGI-S, CGI-Improvement (CGI-I), and CGI-Efficacy. The study physician conducted all the CGI evaluations. There were two types of CGI evaluations. One evaluation focused on GI symptoms (CGI-GI), and one focused on PTHS-related symptoms (CGI-PH) not including GI symptoms. The rationale was that we hoped that MTT would result in substantial improvements in GI symptoms, but it was unclear if it would have any effect on PTHS symptoms. Since GI symptoms are distinct from PTHS symptoms, we thought it best to separate the two by using two different assessments by the study physician. CGI Severity is rated on a 7-point scale: 1. Normal, not at all ill 2. Borderline mentally ill 3. Mildly ill 4. Moderately ill 5. Markedly ill 6. Severely ill 7. Among the most extremely ill patients. The improvement scale uses the following definitions 1. Very much improved 2. Much improved 3. Minimally improved 4. No change 5. Minimally worse 6. Much worse 7. Very much worse.

Parent Global Impressions – Pitt Hopkins syndrome (PGI-PTHS)

There is no validated or unvalidated instrument for assessing symptoms in PTHS patients, as this is the first treatment study for PTHS. Therefore, we adapted the Parent Global Impressions of Autism – Revised (PGI-R3) for patients with PTHS. The following 9 items that are common in PTHS patients, based on a discussion with the Pitt Hopkins Foundation:

1. Hyperventilation
2. Daytime Apnea
3. Gross Motor Skills
4. Fine Motor Skills
5. Ataxia (lack of communication between brain and body)
6. Food allergies
7. Environmental allergies
8. Flushing
9. Rashes

An initial severity score is calculated for each item on a scale of 0 (no symptom) to 6 (extremely severe). At end of Part 1, 2, and 3, a Likert scale is used to rate the change in severity compared to baseline, from much worse (-3) to no change (0) to slightly better (+1), better (+2), and much better (+3). Expressive Language, Receptive Language, Play Skills, Cognition/thinking, Attention/focus, Stool/GI Problems, Sleep, Sociability, Eye Contact, Hyperactivity, Tantrums/meltdowns, Irritability/mood, Anxiety Stimming/perseveration, Sensory sensitivity, Aggression, Self-abuse, Seizures, Self-limited diet, Hyperventilation, Daytime apnea, Gross Motor ,Fine Motor, Ataxia, Food allergies, Environmental allergies, Flushing, Rashes.

Gastrointestinal Symptom Rating Scale (GSRS)

The GSRS is an assessment of GI symptoms based on 15 questions, which are then scored in five domains: abdominal pain, reflux, indigestion, diarrhea, and constipation. We used the GSRS-Likert, which rates each symptom on a 7-point scale. The GSRS is somewhat limited in that the scale is qualitative, but we think it provides good coverage of GI symptoms common in PTHS patients. Since it is not validated for PTHS patients, we used it as a secondary outcome. The 7-point scale: (1) No discomfort at all. (2) Slight discomfort. (3) Mild discomfort. (4) Moderate discomfort. (5) Moderately severe discomfort. (6) Severe discomfort. (7) Very severe discomfort.

Revised Face Legs Activity Crying Consolability Pain Questionnaire for Children with Cognitive Impairment (FLACC)

The original FLACC was slightly revised for children with cognitive impairment (Malviya et al., 2006) by modifying the descriptions of some items. It assesses pain in five areas (Face Legs Activity Crying Consolability) using a three-point scale for each area, resulting in total FLACC pain scale ranging from zero to 10. The Interrater reliability has high intraclass correlation coefficients (ICC, ranging from 0.76 to 0.90) and adequate kappa statistics (0.44-0.57). The criterion validity is supported by good correlations between FLACC, parent, and child scores (rho = 0.65-0.87; $P < 0.001$). The construct validity was demonstrated by a significant decrease in FLACC scores following analgesic administration (6.1 +/- 2.6 vs 1.9 +/- 2.7; $P < 0.001$) in a small study. So, we believe it will be useful for assessing pain in children with PTHS.

Occupational Therapy (OT) Evaluation

The study coordinators conducted telehealth discussions with the study participants to complete the Saltin Grimby Scale and the Sensory Processing Scale.

For the assessment, the DSR, GSRS, PGI-PTHS, FLACC, Vancomycin AE form, and MT AE form were completed by the patient's primary caregiver on an online survey form. The CGI was completed by the study physician after interviewing the patient's primary caregiver and scores were collected on paper.

Randomization scheme

Participants were randomized by research biostatistician in a 1:1 ratio to active or placebo treatment groups using permuted block randomization using function *blockrand* implemented in R software (Snow, 2013). Following enrollment and prescreening, allocation was implemented by an unblinded study

pharmacist, who was responsible for assigning and dispensing study product. Active treatment and placebo were prepared in sequentially numbered containers and uniformly labeled as "medication/placebo" to maintain allocation concealment; the products were visually identical. Participants and study coordinators remained blinded through completion of Part 1 of the study. Study physicians remained blinded through Part 2 and were unblinded after completion of Part 2, whereas the principal investigator remained blinded until completion of Part 3. The study pharmacist remained unblinded throughout the study. Blinding was maintained by using identical labels on active treatment and placebo. The pharmacist kept the lot number in a secure log to preserve allocation concealment. The bowel cleanse was not blinded because its effects were obvious; however, all participants completed the bowel cleanse using the same procedure.

Reference: Snow G. blockrand: Randomization for block random clinical trials. R package version 1.5. 2013. https://cran.r-project.org/web/packages/blockrand/index.html

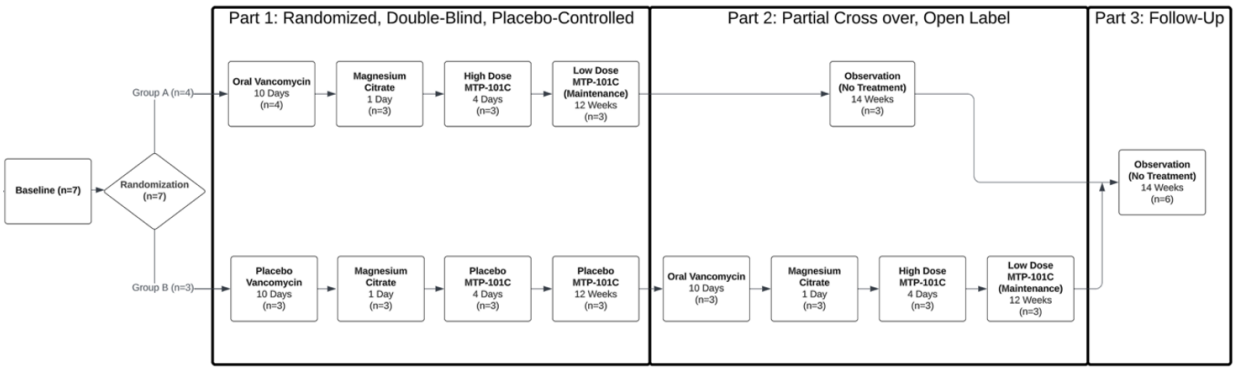

Figure S1. Detailed overview of the study design of randomized double-blind placebo-controlled study in children with PTHS.

Sample size justification:
PTHS is an ultra-rare neurodevelopmental disorder affecting approximately 1,000 individuals in the United States, with chronic gastrointestinal symptoms reported in approximately 70% of affected individuals. Owing to the rarity of the condition and limited participant availability, this pilot Phase II study was designed to enroll 10 participants to obtain initial estimates of safety, tolerability, dosage, and efficacy of microbiota transplant therapy (MTT).

Safety analysis: Safety was evaluated based on the frequency and severity of adverse events and serious adverse events, as well as changes in blood safety parameters including blood chemistry panels and complete blood counts with differential.

Tolerability analysis: Tolerability was assessed based on study medication adherence, measured by percentage of medication consumed, and participant retention throughout the study.

Efficacy analysis: The primary efficacy outcome was change in Daily Stool Record (DSR) scores. Secondary efficacy outcomes included changes in Clinical Global Impressions (CGI), Parent Global Impressions–Pitt Hopkins (PGI-PTHS), and Gastrointestinal Symptom Rating Scale (GSRS) scores.

Informed consent: Written informed consent was obtained from parents or legal guardians before study participation. When developmentally appropriate, participants were also provided an age-adapted explanation of the study and given the opportunity to ask questions and provide assent.

Compensation: Study medications were provided at no cost to participants. No financial compensation was provided for study participation.

Changes in the protocol: No major changes were made to the trial protocol after study commencement, and no outcomes or analyses were added that were not prespecified. The only modification involved the method of data collection. Initially, study data were to be collected using the I Am Not Alone (IMNA) Solutions digital health software platform. However, due to technical difficulties with the IMNA software, its use was discontinued in early phase of the trial. Prior to discontinuation, study coordinators transferred all previously collected data (including part of the baseline data) from the IMNA system. Thereafter, all study data were collected using paper forms and/or through direct communication with study coordinators, in accordance with the original study protocol.

Interim analysis: No prespecified interim statistical analyses or stopping boundaries for efficacy were implemented. Study data were analyzed after all participants completed the protocol and follow-up assessments. Safety monitoring was conducted throughout the study, and annual reports were submitted to the FDA in October 2023 and October 2024.

Statistical/Analytical Issues

Overall, patient compliance with the study protocol was excellent, with assessments and forms rarely missed. The missing data included:

Daily Stool Record: Two patients in Group A did not report days of pain at baseline. They did report substantial improvements in abdominal pain on the GSRS, so if their DSR data had been reported it likely would have increased the Cohen's D effect size for the DSR.

Sensory Processing Form: The baseline occupational therapy evaluation data were unavailable for two Group B patients due to loss of data during COVID transition, and one Group A patient did not complete the end of Part 3 questionnaire on occupational therapy.

Adjustment for Covariates:

There were no adjustments for covariates, due to small sample size.

Handling of Dropouts or Missing Data:

One patient was withdrawn from the study on Day 4 due to being unable to consistently swallow capsules. This patient only contributed data at baseline and for a few days into vancomycin/placebo treatment during Part 1 of the study. All of their data was excluded from analysis.

In cases where specific data points were missing, analyses were conducted using the available data, and the sample size for each analysis that was missing data was reported accordingly.

Interim Analyses and Data Monitoring:

Study data was monitored on an ongoing basis by the study coordinators to ensure accuracy, completeness, and protocol compliance. In addition, patients met with a coordinator and the Principal Investigator at baseline, end of Part 1, end of Part 2, and end of Part 3 to discuss the survey responses and allow for clarification and verification of reported data.

Data was analyzed once all patients completed the study. A report describing the results was sent to the FDA in the annual report on October 2023 and October 2024.

Multicenter Studies:

This study was conducted at a single site, with patients remaining in their home state throughout the study. Patients used their local PCP for physical exams and vital monitoring and their local Lab for routine blood safety labs. Study supplies were mailed directly to patients, while patient samples were

returned to Arizona State University's Biodesign Center for appropriate storage and processing. Medications were shipped to patients from Pure Compounding Pharmacy in Chicago.

Multiple Comparisons/Multiplicity:
Given the small size of the study, no corrections for multiple hypothesis testing were applied. Each analysis was treated as an independent assessment.

Use of an "Efficacy Subset" of Patients:
There was no use of an efficacy subset of patients.

Active-Control Studies Intended to Show Equivalence and *Examination of Subgroups* were not applicable.

Diet

Table S1: Diet overview throughout the study

Study ID	Maintained same diet throughout the study	Any major food types or dietary patterns	Any dietary changes or comments about food intake during the treatment period
PH-006	Yes	None reported	None reported
PH-009	Yes	None reported in Part 1, 2 and 3 except->	**Part 1-** reported a slight loss of appetite during the last 7 days of the Vanco phase, which resolved upon starting the microbiota phase
PH-002	Yes	None reported in Part 1, 2 and 3 except->	**Part 2**- slight appetite decrease that resolved within 1 day during the first days of treatment.
PH-003	Yes	None reported	None reported
PH-004	Yes	None reported	None reported
PH-005	Yes	None reported	**Part 3** - Nutritional shake was stopped for 8 weeks due to recall on 2/17/2022. Resumed after this period.

Mitochondrial function test:

Initially, from each participant, Buccal cells were harvested using sterile foam tipped applicators/swab (cat# 25-1506-1PF 100, Puritan Medical Products Company, Maine, USA) by firmly pressing a swab against the inner cheek while twirling for 30 s on each cheek. Typically, 2 swabs from each participant were used for extraction of buccal mitochondrial protein. Buccal swab samples were collected at baseline, end of Part 1, and end of Part 2 for assessment of mitochondrial function. Buccal samples from typically developing (TD) children of comparable age (12.7 ± 2.9 years vs. 12.2 ± 3.1 years) and sex (4 males, 2 females vs. 6 males, 1 female) served as controls. Samples were shipped to *Religen Inc.* (https://religendx.com), a CLIA-certified laboratory, for analysis using MitoSwab® assay, which measures mitochondrial respiratory chain enzyme activities (Complex IV, Cytochrome-c Oxidase) (Goldenthal et al., 2011). Each swab was clipped and then placed in a 1.5 ml microcentrifuge tube with 120-150 µl of ice cold extraction buffer A containing 1.5% lauryl maltoside, 100 mM of NaCl, 25 mM of HEPES pH 7.4, as well as a protease inhibitor cocktail (Sigma, P-8340). After incubation for 20 min on ice, the swab-containing tubes were vortexed for 20 s and then cold centrifuged at 8200 ×g for 5 min to remove the extracted protein from the swabs. The sample extracts were cleared

of insoluble material by centrifugation at 8200 ×g for 8 min at 4 °C. Duplicate aliquots were analyzed for protein concentration determination using the Bicinchoninic acid method (Pierce, USA). Samples were stored as single use aliquots at −80 °C for further enzymatic analysis. Total protein concentration was determined by Bicinchoninic acid (BCA) method (Pierce, USA) using bovine serum albumin as a standard.

Complex IV (cytochrome-c oxidase) activity was quantified in duplicate by monitoring the oxidation of reduced cytochrome-c at 550 nm using a standard spectrophotometric procedure in 0.5 ml reaction volume, in HITACHI (USA) U2910 UV-VIS Spectrophotometer. Specific activities of RC I and RC IV were normalized relative to buccal CS specific activity levels, a commonly used gauge of overall mitochondrial content and were expressed as activity ratios (i.e. IV/CS). The use of activity ratios as compared to absolute activities in mitochondrial enzyme assessment is well established [1, 2, 3,] and provides a much narrower range of normal control values as compared to activities expressed on the basis of protein content of the sample, particularly critical for buccal analysis.

Assays were performed in duplicate/triplicate (technical/biological) following Religen's standard quality-control procedures. For CS and RC-IV enzyme assays, which are kinetic assays, the molar extinction coefficient is used to calculate the specific activity. For quality control, 3 prepared standards (STDs) at High, Medium and Low levels over the normal range were stored in single use aliquots and are used along with every assay before the samples are assayed. For RC-I 5 levels of STDs covering the standard operating range of values were used in similar way. Enzyme activity was reported as nanomoles of cytochrome-c oxidized per minute per milligram of protein (nmol min^{-1} mg^{-1} protein).

Symptoms results:

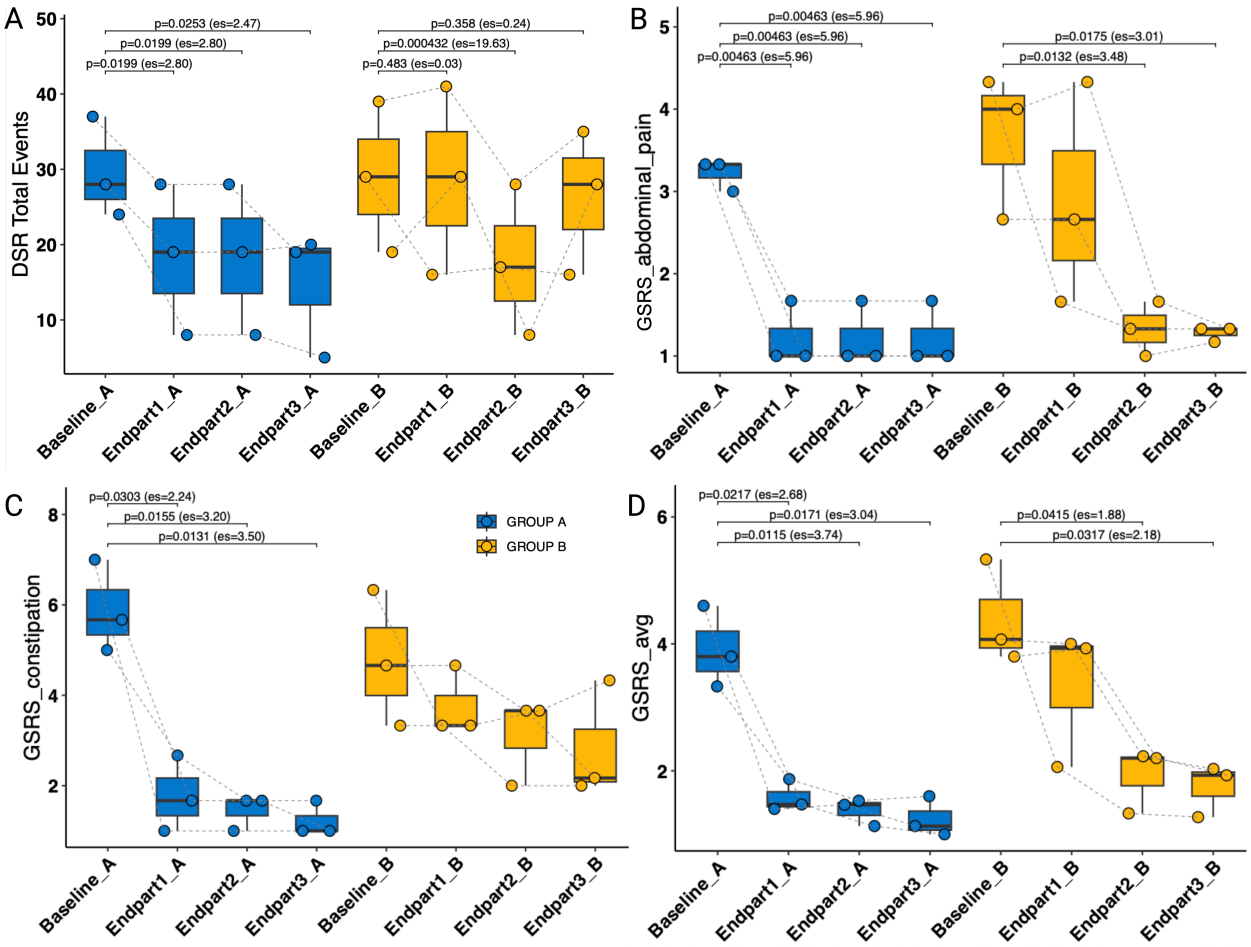

Figure S2: Longitudinal changes in Daily Stool Record (DSR), Gastrointestinal symptoms before and after MTT in children with PTHS. Figure show the, A) DSR-Total events; B) GSRS-Abdominal pain; C) GSRS-Constipation; D) GSRS-average in group-A and -B. A paired t-test (one-tailed) was used to compare between timepoints, with $p<0.05$ considered statistically significant with Cohen's *d* effect size are displayed on the plot. A=group-A; B=group-B; GSRS= Gastrointestinal Symptom Rating Scale; es=effect size. Between=Baseline_B and EndPart3_B.

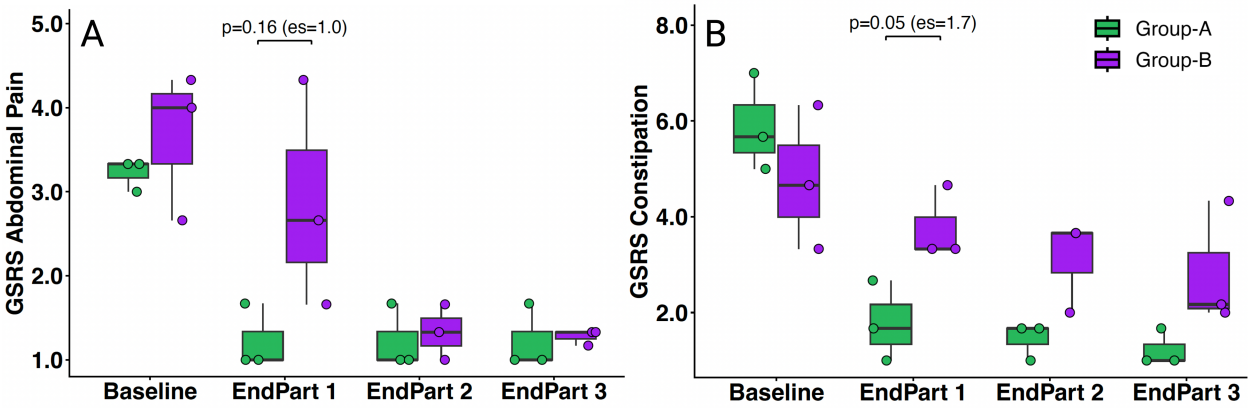

Figure S3: PTHS related GSRS symptoms before and after MTT in children with PTHS. Changes in A) GSRS-Abdominal pain; B) GSRS-Constipation. A paired t-test (one-tailed) was used to compare between timepoints, with $p<0.05$ considered statistically significant with Δ Cohen's *d* effect size are displayed on the plot. Absolute symptoms scores were used to make the plots and delta values between baseline and different time points were used for the statistical analysis. A=Group-A; B=Group-B; DSR=Daily Stool Records GSRS= Gastrointestinal Symptom Rating Scale; es=Cohen's *d* effect size. For the longitudinal comparison within Group-A and -B

Primary Outcome:

1. Daily Stool Record (DSR)

The DSR was recorded daily for a 14 day period at baseline, end of Part 1, end of Part 2, and end of Part 3. It includes a recording of the number of stools, the stool type (per Bristol Stool Form Scale), GI medications each day, and whether there was significant GI pain each day.
Below we first discuss the summary score based on Number of Abnormal Days, a new summary score based on Number of Total Events, and then each subscore.

1.1. New Method for Scoring Daily Stool Record: DSR Total Events*:* The original method for the scoring of the Daily Stool Record involved calculating the number of Abnormal Days in which one or more symptoms were present, or if GI medications were required. However, because children with PTHS have severe GI symptoms despite the use of daily GI medications, the scoring of improvements in symptoms was masked in many cases by the continued use of GI medications, so that almost every day counted as an Abnormal Day. We decided that a day with 3 events such as GI medications, pain, and constipation should be scored as more severe than a day with only 1 event such as GI medications. So, we developed a new scoring algorithm, based on simply calculating the Total Events each day for 14 days; i.e., use of GI medications, abdominal pain, no stool, unusually hard stool, unusually soft/liquid stool, and/or 4+ stools/day, with the Bristol Stool Form scale used to identify days of unusually hard (type 1-2) or soft/liquid stool (type 6-7). Using this algorithm, we found the following results:

In Part 1, Group A had a 38% decrease in number of events, vs 1% decrease for Group B (placebo), resulting in a Cohen's d effect size of 1.34 (very large), Table S3

In Part 2, Group A continued to have good improvement 3 months after treatment stopped. Group B improved after they received the unblind real treatment resulting in an estimated effect size of 1.89(very large), Table S3.
In Part 3, Group A improved slightly more at 6 months after treatment stopped. Group B did not maintain improvement in 2 patients; in 1 case this was due to dehydration due to loss of access to a favorite nutritional shake, and in 1 case the patient needed to go back on GI medications. The estimated long term effect size is
0.85 (large), Table S3.

Although this is a post-study analysis, we believe it provides more insight into the original DSR analysis and yields a similar effect size for Part 1.

Table S2: Average Number of DSR Total Events Over a 14 Day Period

	Baseline (Events)	**End Part 1 (Events)**	**End Part 2 (Events)**	**End Part 3 (Events)**
Group A	29.7	18.3	18.3	14.7
Group B	29.0	28.7	17.7	26.3
p-value*		0.11		
Cohen's d		1.34		
Estimated Cohen's d (Group A and B combined)			1.89	0.85
*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.				

Table S2-A: Percent Change of Average Abnormal DSR Total Events Compared to Baseline

	End of Part 1	**End of Part 2**	**End of Part 3**
Group A	-38%	-38%	-51%
Group B	-1%	-39%	-9%

1.2 Number of Abnormal Days: An abnormal day is defined as a day with no stool, an unusually hard stool (type 1-2), unusually soft stool (type 6-7), 4+ stools, significant abnormal pain, or GI medication use. The data was primarily affected by days of GI medication use (very high), and somewhat by days of soft stool and days of significant pain.
For Group A, there was a small reduction in the number of abnormal days in Part 1, and that continued in Parts 2 and 3. For Group B, there was no change in Part 1, a small decrease in Part 2 , and number of abnormal days returned to baseline scores at Part 3(due partially to a patient no longer having access to their nutritional shake and hence having less fluid intake and more constipation), Table S2.
The Cohen's d effect size is 1.3 (very large).

The estimated Cohen's d for end of part 2 is 0.82 (large) and it is 0.48 (Medium) at end of part 3.

Table S3: Average Number of Abnormal Days Over 14 days				
	Baseline (Days)	**End Part 1 (Days)**	**End Part 2 (Days)**	**End Part 3 (Days)**
Group A	14.0	11.7	12.0	11.0
Group B	14.0	14.0	12.0	14.0
p-value*		0.12		
Cohen's d		1.3		
Estimated Cohen's d (Group A and B combined)			0.82	0.48
*t-test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.				

Table S3-A: Percent Change of Average Abnormal Days Over 14 Days Compared to Baseline

	End of Part 1	**End of Part 2**	**End of Part 3**
Group A	-17%	-14%	-21%
Gorup B	0%	-14%	0%

1.3 GI Medication Usage: In this study the DSR is complicated because all patients were taking GI medications at the start of the study, as described below. Patients were allowed to reduce GI medications during the study if they were no longer needed.

The DSR accounts for changes in frequency of medications, but not for changes in dose. In this study patients often reduced both, so reporting only changes in frequency is an under-estimate of the effect. This is a limitation of the current DSR.

At baseline, all patients were taking GI medications to treat constipation.

During Part 1, there was more decrease in usage of GI medications by Group A than by Group B (see summary table below). By the end of Part 2, three patients had large decreases in medication usage, 1 had medium decrease, 1 had small decrease, and 1 had no change. These decreases generally continued until the end of Part 3 except for one patient (005) who lost access to their nutritional shake in the middle of Part 3 and hence had decreased fluid intake and had to start daily MiraLAX to treat constipation

Table S4 shows a summary table of changes in GI medication usage compared to baseline, including changes in dosage and/or frequency

Table S4: Summary of Changes in GI Medication Usage by Patient Over Time

	Part 1	**Part 2**	**Part 3**
Group A			
PH-003	Small decrease	Small decrease	Small decrease
PH-006	Large decrease	Large decrease	Large decrease
PH-009	Large decrease	Large decrease	Large decrease
Group B			

PH-002	Medium decrease	Medium decrease	Medium/large decrease
PH-004	No change	No change	Small decrease
PH-005	Medium decrease	Large decrease	Large increase

The details on medication usage for each patient are given below:

Group A:
Patient PH-003: No longer needed Fleet enema during Parts 1, 2, or 3
Baseline: daily use of Lactulose and MiraLAX; 1 Fleet enema
Parts 1, 2, 3: daily use of lactulose and MiraLAX, but no Fleet enemas
Patient PH-006: Reduced their frequency of MiraLAX by approximately 50%, and reduced the dosage by over 50%.
Baseline: 12 days of MiraLAX (2 days of 4 tbsp, 10 days of 2 tbsp)
Part 1: 6 days of MiraLAX (1 day of 2 tbsp, 4 days of 1 tbsp) and 1 Fleet enema
Part 2: 7 days of MiraLAX (1 tbsp)
Part 3: 5 days of MiraLAX (1 tbsp)
Patient PH-009: Stopped their daily use of Benefiber in Part 1 and Part 2
Baseline: daily use of Benefiber (14 days)
Part 1 and 2: no Benefiber or other GI meds
Part 3: 1 day of Benefiber

Group B:
Patient PH-002: Decreased their dosage of Senna by 50% in Parts 1, 2, and 3; decreased need for enema in Parts 2 and 3.
Baseline: daily use of Senna (2x/day, 2 pills/dose) and 2 enemas
Part 1: daily use of Senna (1x/day, 2 pills/dose) and 2 enemas
Part 2: daily use of Senna (1x/day, 2 pills/dosa) and 1 enema
Part 3: daily use of Senna (1x/day, 2 pills/dose); no enema
Patient PH-004: In Part 3 there was a partial reduction of dosage of magnesium oxide
Baseline, Part 1, Part 2: daily Senna and magnesium oxide; 4 days of extra of magnesium oxide
Part 3: daily Senna and magnesium oxide (no days of extra magnesium oxide
Patient PH-005: Decreased suppository usage in Part 1, and more decrease in Part 2, but added MiraLAX in Part 3.
Baseline: 2 days of suppository
Part 1: 1 day of suppository
Part 2: no suppository
Part 3: daily use of MiraLAX (halfway into Part 3 the patients nutritional shake became unavailable, so they consumed less fluid and became constipated).

During Part 1, the treatment Group A reduced the number of days taking GI medications, but the placebo Group B did not. This resulted in a Cohen's D of 1.27 (very large) **Table S4**.

In Part 2, Group A continued at a lower number of days taking GI medications, and that continued into Part 3. Group B stayed at the same amount of GI medications in Part 2; but in Part 3 one patient changed from 1-2 GI medications per 14 days to daily GI medications

because their nutritional shake became unavailable, which resulted in less fluid intake and more constipation**, Table S4**. This resulted in a medium effect size in part 2 (Estimated Cohen's d=0.73) and a very small effect in part 3 (Estimated Cohen's d = 0.14).

Table S4: Average Number of Days of Medication Use Over 14 Days

	Baseline (Days)	**End Part 1 (Days)**	**End Part 2 (Days)**	**End Part 3 (Days)**
Group A	13.3	6.7	7.0	6.7
Group B	10.0	9.7	9.3	14.0
p-value*		0.13		
Cohen's d		1.27		
Estimated Cohen's d (Group A and B combined)			0.73	0.14

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part1

1.5 Days of No Stool: The number of days of no stool (out of 14 days) was generally low because patients were taking medications to treat constipation.

In Part 1, Group A had a small increase (worsening), and Group B had a small decrease (improvement). The effect size was large (Cohen's d= -1.31) **Table S5**.

In Parts 2 and 3, the number of days of no stool was similar to the end of Part 1 and overall Group B showed more improvement than Group A at both timepoints. The effect size was large for Part 2 (Estimated Cohen's d=-0.84) and medium for Part 3 (Estimated Cohen's d= -0.71), **Table S5.**

Table S5: Average Number of Days No Stool Over 14 Days

	Baseline (Days)	**End Part 1 (Days)**	**End Part 2 (Days)**	**End Part 3 (Days)**
Group A	3.3	4.3	4.3	4.0
Group B	3.7	2.3	3.0	3.0
p-value*		0.11		
Cohen's d		-1.3**		
Estimated Cohen's d (Group A and B combined)			-0.84**	-0.71**

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.
**Placebo group improved more.

1.6 Days of Hard Stool (Bristol Stool Form type 1-2: The number of days of hard stool remained very low throughout the study, partially because of the GI medications that were used to keep the stool soft so it could be passed. Note that decreases in usage of GI medication did not worsen days of hard stool.

In Part 1, both Groups A and B had a small increase in the number of days of hard stool, with Group A increasing slightly more. The effect size was small (Cohen's d=0.47) **Table S6**.

In Part 2, both Groups A and B had a small decrease in the number of days of hard stool. The estimated effect size is very small (Cohen's d=-0.29) **Table S6**.

In Part 3, Group A continued to have a small number of days of hard stool, and Group B had a small increase. The effect size was very small (Cohen's d=0.07) **Table S6**.

Overall, the number of days of hard stool remained low during the study with small fluctuations.
Table S6: Average Number of Days of Hard Stool over 14 Days

	Baseline (Days)	**End Part 1 (Days)**	**End Part 2 (Days)**	**End Part 3 (Days)**
Group A	0.2	1.3	0.7	0.3
Group B	0.3	0.7	0.0	1.0
p-value*		0.31		
Cohen's d		0.47		
Estimated Cohen's d (Group A and B combined)			-0.29**	0.07

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.
**Placebo group improved more.

1.7 Days of Soft Stool (Bristol Stool Form type 6-7): At baseline, patients averaged 8 days out of 14 with very soft or liquid stool.

In Part 1, there was a small decrease in the number of days with soft stool for both Groups. The effect size was very small (Cohen's d=0.03) **Table S7**.

In Part 2, there was an additional decrease when Group B was switched to the unblind treatment. The estimated effect size was very small (Estimated Cohen's d=0.19) **Table S7**.

In part 3, Group A had a further decrease in number of days of soft stool. Overall, combining Groups A and B, by the end of Part 3 there was a 45% decrease in the number of days of soft stool compared to baseline. The estimated effect size was small (Estimated Cohen's d=0.44) **Table S7**.

Table S7: Average Number of Days of Soft Stool Over 14 Days

	Baseline (Days)	**End Part 1 (Days)**	**End Part 2 (Days)**	**End Part 3 (Days)**
Group A	7.8	6.0	6.3	3.3
Group B	8.7	7.0	5.0	5.7
p-value*		0.49		
Cohen's d		0.03		

Estimated Cohen's d (Group A and B combined)			0.19	0.44

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.

1.8 Days of 4 or More Stools per Day: The patients had very few days of 4 or more stools per day, and that continued to be the case throughout the study.

In Part 1, both Groups decreased the number of days of 4+ stools, with a slightly greater decrease for Group A. The effect size was small (Cohen's d=0.37). Table S8.

In Part 2, Group A stayed at zero days, and Group B decreased to zero days. The estimated effect size was small (Estimated Cohen's d=0.46). Table S8.

In Part 3, both Groups remained at near-zero days of 4+ stools. The estimated effect size was small (Estimated Cohen's d=0.31). Table S8

Table S8: Average Days of 4 or More Stools Over 14 Days

	Baseline (Days)	**End Part 1 (Days)**	**End Part 2 (Days)**	**End Part 3 (Days)**
Group A	0.7	0.0	0.0	0.3
Group B	1.0	0.7	0.0	0.0
P-value*		0.34		
Cohens d		0.37		
Estimated Cohen's d (Group A and B combined)			0.46	0.31

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.

1.9 Days of Significant Pain: In Group A, the parents of two patients in Group A did not feel that they could accurately rate their child's pain on a daily basis, so they did not record it. However, on the weekly GSRS they reported an average 83% reduction in abdominal pain at end of Part 1, and 100% at end of Part 2 and end of Part 3, so their results are similar to the results reported below for 1 patient in Group A who did have data.

In Part 1, the patient in Group A had a 100% reduction in days of significant pain; the 3 patients in Group B reported an increase in days of significant pain Table S9.

In Parts 2 and 3, the patient in Group A continued to have no days of significant pain. The patients in Group B reported a major decrease in days of significant pain in Part 2 (after treatment), and the days of pain in Part 3 was almost as low Table S9.

Overall, by the end of Part 2, 4 patients had a 97% decrease in the number of days of significant pain, and a 72% decrease at end of Part 3 Table S9.

At the end of Part 3, a long-term effect size can be estimated by combining Groups A and B at end of Part 3 and comparing them vs placebo (Group B at end of Part 1). We use the phrase "estimated" because we are comparing open-label treatment (end of Part 3) vs placebo at end of Part 1. This yields an estimated long-term effect size (Cohen's d) of 1.25 ("very large", defined as 1.2-1.99) Table S9.

Table S9: Average Number of Days of Pain over 14 Days

	Baseline (Days)	**End Part 1 (Days)**	**End Part 2 (Days)**	**End Part 3 (Days)**
Group A	13.0	0.0	0.0	0.0
Group B	5.3	8.3	0.3	2.7
p-value*		*		
Cohen's d		*		
Estimated Cohen's d (Group A and B combined)			*	*

*T-test, Cohen's d and estimated Cohen's d could not be calculated because of missing data for two patients in Group A

Summary of DSR: The DSR was useful in evaluating many aspects of GI symptoms in patients with PTHS. However, the original scoring method of Number of Abnormal Days was limited because patients often had multiple events on a single day, so everyone had a score of 14 abnormal days out of 14 at baseline. Therefore, we introduced the Total Events score to better capture the severity of GI disorders in PTHS.

MTT resulted in improvements in the DSR. For Part 1, the calculated Cohen's d effect size for Number of Abnormal Days was 1.3 (very large), but the percent change was small (17% for Group A vs 0% for Group B) because patients had multiple symptoms which resulted in almost all days remaining abnormal even though individual symptoms improved. Re-analysis with the new method of Total Events yielded a similar effect size (1.3), but a larger percent change for Group A (-37% for Group A vs -1% for Group B). In Part 2, Group A retained improvements at 3 months post-treatment, and Group B was switched to treatment and also improved. In Part 3, Group A retained their original improvements, but Group B lost benefit at 3 months post-treatment, partially due to one patient losing access to their nutritional shake, and one patient needing to increase their GI medications.

Regarding the DSR subscores, the major symptoms at baseline were medication use, significant pain, and soft stools. The largest improvement in the DSR was the number of days of significant pain (72% decrease by end of Part 3 averaging over the 4 patients who reported data). There was some decrease in number of days of soft stool (43% decrease by end of Part 3 averaging over all patients). There was some decrease in both medication frequency and medication dosage. There was some decrease in the number of days of medication usage in Part 2 (30%) but less at end of Part 3 (11%). By the end of Part 2 many patients had decreased their medication dosage (3 large decreases, 1 medium decrease, 1 small decrease, and 1 no change) and this generally continued

until the end of Part 3 except for 1 patient who added daily MiraLAX due to loss of their nutritional shake. The other DSR symptoms (days of no stool, days of hard stool, days of 4+ stool) were much less common at baseline (less than 1.5/day) and did not change much during the study.

The reported improvement on the DSR is under-estimated because:

1) 2 patients in Group A did not score abdominal pain on a daily basis, but did report large improvements in abdominal pain on the GSRS.
2) The DSR captured the reduction in medication frequency, but was not designed to capture reductions in dosage which also occurred.

Thus, the actual improvement in GI symptoms was slightly under-estimated by the DSR.

Secondary Outcomes:

2. Gastrointestinal Symptom Rating Scale (GSRS) – Likert version:

The GSRS-Likert consists of 15 items, rated on a scale of 1 (No discomfort at all) to 7 (Very severe discomfort). There are five subscales and an average score of all 15 items. Note that the GSRS-Likert uses simpler lay language than the original GSRS since it is intended for patients, and it uses a 7-point scale of severity instead of the original 4-point scale.

1) No discomfort at all.
2) Slight discomfort.
3) Mild discomfort.
4) Moderate discomfort.
5) Moderately severe discomfort.
6) Severe discomfort.
7) Very severe discomfort.

2.1 GSRS-Average of all 15 items: At baseline, averaging over all 15 symptoms, Group A had moderate symptoms and Group B had moderate to moderately severe symptoms.

At the end of Part 1, Group A had much more reduction in GSRS average score (-80%) than Group B (-31%). The effect size is large (Cohen's d=0.85) **Table S10**.

In Part 2, Group A retained their improvements in GI symptoms at 3 months post-treatment. Group B also improved substantially when they received the unblind treatment. The estimated effect size is large (Estimated Coehn's d=1.11) **Table S10**.

In Part 3, the improvements remained at the end of Part 3 for both groups. Combining both Groups, at the end of Part 3 compared to baseline there was an 84% reduction in GSRS average score (p=0.001 based on 2-sided paired t-test). The estimated Cohen's d is very large
(Estimated Cohen's d=1.23) **Table S10, Table S11**.

Table S10: GSRS (Average Severity of 15 Items)

	Baseline (Severity)	**End Part 1 (Severity)**	**End Part 2 (Severity)**	**End Part 3 (Severity)**

Group A	3.9	1.6	1.4	1.2
Group B	4.4	3.3	1.9	1.7
p-value*		0.19		
Cohen's d		0.85		
Estimated Cohen's d (Group A and B combined)			1.11	1.23

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.

Table S11: Percent Change in GSRS Average Score Compared to Baseline

	End Part 1	**End Part 2**	**End Part 3**
Group A	-80%	-87%	-92%
Group B	-31%	-73%	-78%

2.2 GSRS- abdominal pain subscale: The GSRS-abdominal pain subscale is an average of three items: abdominal pain, hunger pains and nausea. At baseline, both Groups averaged in the mild to moderate range of severity Table S12.

By the end of Part 1, Group A improved much more than Group B on the abdominal pain subscale. The effect size is large (Cohen's d=1.02) Table S12.

At the end of Part 2, both Groups had greatly improved, and those improvements lasted until the end of Part 3. The estimated effect size is very large for end of part 2 and 3 (Estimated Cohen's d=1.42 and 1.4, respectively) Table S12.

Table S12: GSRS Average Score of Abdominal Pain Subscale

	Baseline (Severity)	**End Part 1 (Severity)**	**End Part 2 (Severity)**	**End Part 3 (Severity)**
Group A	3.2	1.2	1.2	1.2
Group B	3.7	2.9	1.3	1.3
p-value*		0.16		
Cohen's d		1.02		
Estimated Cohen's d (Group A and B combined)			1.42	1.4

2.3 GSRS Reflux Subscale: The GSRS reflux subscale is an average of two items, heartburn and acid regurgitation.

Group A had minimal reflux symptoms, and they remained minimal throughout the

study, Table S13.

Group B had moderate reflux symptoms at baseline, and had some decrease in reflux at end of Part 1, and near zero symptoms at the end of Part 2, and no symptoms at the end of Part 3, Table S13.

Table S13: GSRS – Average Score of Reflux Subscale

	Baseline (Severity)	**End Part 1 (Severity)**	**End Part 2 (Severity)**	**End Part 3 (Severity)**
Group A	1.3	1.0	1.0	1.2
Group B	4.2	2.8	1.2	1.0
p-value		*		
Cohen's d		*		
Estimated Cohen's d (Group A and B combined)			*	*

*The t-test, Cohen's D and estimated Cohen's D were not calculated because Group A had near-zero symptoms at baseline, whereas Group B had substantial symptoms.

2.4 GSRS Diarrhea Subscale: The GSRS diarrhea subscale is an average of three items: diarrhea, loose stools and urgent need for defecation. At baseline, both groups had moderate severity **Table S14**.

At the end of Part 1, Group A had substantially more decrease in diarrhea subscale than Group B. The effect size was of medium size (Cohen's d=0.70) **Table S14**.

At the end of Part 2, Group B had more decrease in diarrhea. The improvements for Group A and B continued until the end of Part 3. The estimated effect size was medium for end of part 2 and end of part 3 (Estimated Cohen's d=0.56 and 0.61, respectively) **Table S14**.

Table S14: GSRS Average Score of Diarrhea Subscale

	Baseline (Severity)	**End Part 1 (Severity)**	**End Part 2 (Severity)**	**End Part 3 (Severity)**
Group A	4.0	1.1	1.1	1.1
Group B	4.2	3.0	1.8	1.6
p-value*		0.23		
Cohen's d		0.70		
Estimated Cohen's d (Group A and B combined)			0.56	0.61

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.

2.5 GSRS Indigestion Subscale: The GI indigestion subscale is an average of four items:

stomach rumbling, bloating, burping and gas.

At baseline, Group A had moderate symptoms, and Group B had moderate to moderately severe symptoms Table 23.

At the end of Part 1, Group A improved substantially more than Group B. The effect size is moderate (Cohen's d=0.74) **Table S15**.

At the end of Part 2, Group A improved slightly more, and Group B improved substantially. The estimated effect size is large (Cohen's d=1.08) **Table S15**.

At the end of Part 3, both Groups remained substantially improved. The estimated effect size is very large (Cohen's d=1.55) **Table S15**.

Table S15: GSRS Average Score of Indigestion Subscale

	Baseline	**End Part 1**	**End Part 2**	**End Part 3**
Group A	4.2	2.3	1.8	1.4
Group B	4.8	3.8	2.0	1.8
p-value*		0.21		
Cohen's d		0.74		
Estimated Cohen's d (Group A and B combined)			1.08	1.55

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.

2.6 GSRS – constipation subscale: The GSRS subscale is an average of three items: constipation, hard stools and incomplete evacuation.

At baseline, Group A had severe symptoms and Group B had moderately severe symptoms, **Table 16**.

At the end of Part 1, Group A improved substantially more than Group B (p=0.05). The effect size was very large (Cohen's d=1.74) **Table S16**.

At the end of Part 2, Group A remained improved at 3 months post-treatment, and Group B improved somewhat more after being switched to the unblind treatment. The estimated effect size for end of part 2 is large (estimated Cohen's d=0.95) **Table S16**.

At the end of Part 3, the improvements remained similar to the end of Part 2. The estimated effect size is very large (estimated Cohen's d=1.02) at end of part 3 **Table S16**.

Table S16: GSRS Average Score of Constipation Subscale

	Baseline (Severity)	End Part 1 (Severity)	End Part 2 (Severity)	End Part 3 (Severity)
Group A	5.9	1.8	1.4	1.2
Group B	4.8	3.8	3.1	2.8
p-value*		0.05		
Cohen's d		1.74		
Estimated Cohen's d (Group A and B combined)			0.95	1.02

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.

Summary of GSRS: In summary, there were good improvements on the GSRS and on each of the subscores, and generally the treatment group improved more than the placebo group in Part 1, and generally both Groups had improved by Part 2, and the improvements generally lasted to the end of Part 3 (3-6 months after treatment stopped).

Interpretation of GSRS vs DSR: It is somewhat puzzling that the patients reported more improvement on the GSRS than on the DSR. However, based on extensive multiple discussions with each patient during the study, and review of the data, we offer the following interpretation. The patients were all on substantial GI medications to treat underlying constipation, which resulted in bowel movements almost every day. However, the GSRS data shows that at baseline the patients had moderately severe or severe incomplete evacuation, which likely caused the moderate to severe bloating reported at baseline, and probably caused much of their abdominal pain and probably contributed to reflux. After treatment, the incomplete evacuation and bloating was much reduced, and pain was much reduced. So, we believe that explains why the patients reported much less pain and improvement in GI symptoms after treatment.

The detailed data on incomplete evacuation and bloating are shown below.

2.7 GSRS – Single Question on Incomplete Evacuation: At baseline, both groups had moderately severe to severe incomplete evacuation.

Part 1: Group A improved much more than Group B (-93% vs -15%, p=0.016). The effect size is huge (Cohen's d=2.71) Table S17.

Part 2: Group A remained improved, and Group B improved somewhat more. The effect size is very large (Estimated Cohen's d=1.24) Table S17.

Part 3: Group A and B remained improved. The effect size is very large (Estimated Cohen's d=1.5) Table S17

Table S17: SRS Average
Score of Incomplete Evacuation

	Baseline (Severity)	End Part 1 (Severity)	End Part 2 (Severity)	End Part 3 (Severity)
Group A	5.7	1.3	1.7	1.3
Group B	5.3	4.7	3.3	3.0
p-value*		0.016		
Cohen's d		2.71		
Estimated Cohen's d (Group A and B combined)			1.24	1.50

.*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.

2.8 GSRS – Single Question on Bloating: At baseline, Group A averaged moderate bloating, and Group B averaged moderately severe to severe bloating. Part 1: Group A improved more than Group B (-67% vs -29%). The effect size is small (Cohen's d=0.35)

Part 2: Group A's symptoms remained improved. Group B improved more. The estimated effect size is medium of size (Estimated Cohen's d=0.69) **Table S18.**

Part 3: Group A's symptoms remained improved. Group B improved more. The estimated effect size is large (Estimated Cohen's d=1.1)**, Table S18.**

Table S18: GSRS Average Score of Bloating

	Baseline (Severity)	End Part 1 (Severity)	End Part 2 (Severity)	End Part 3 (Severity)
Group A	4.0	2.0	1.7	1.3
Group B	5.7	4.3	2.3	2.0
p-value*		0.35		
Cohen's d		0.35		
Estimated Cohen's d (Group A and B combined)			0.69	1.1

* T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.Thus, we believe that the substantial decrease in incomplete evacuation led to a substantial decrease in bloating, which in turn resulted in a substantial reduction in pain.

3. Revised Face Legs Activity Crying Consolability Pain Questionnaire for Children with Cognitive Impairment (FLACC)

The Face-Legs-Activity-Crying-Consolability (FLACC) questionnaire assesses pain on a scale of 0-2 for each of the 5 items, and the scores for each item form the total score (scale of 0-10). The scale is designed for parents assessing children with minimal communication ability.

At the end of Part 1, Group A reported more reduction in pain than Group B. The effect size

was medium of size (Cohen's d=0.56) **Table S19.**

At end of Part 2, Group B reported additional reduction of pain. The estimated effect size was medium of size (Estimated Cohen's d=0.71) **Table S19.**

At the end of Part 3, Group A continued to have little pain, and Group B had slightly more reduction in pain. The estimated effect size was large of size (Estimated Cohen's d=0.99) **Table S19.**

Combining both groups, at the end of part 3 compared to baseline there was a 79% reduction in the FLACC score, p=0.002 (2-sided, paired t-test).

At end of Part 1, the calculated effect size (Cohen's d) was 0.56 ("medium")

At the end of Part 3, a long-term effect size can be estimated by combining Groups 1 and 2 at end of Part 3 and comparing them vs placebo (Group B at end of Part 1). We use the phrase "estimated" because we are comparing open-label treatment (end of Part 3) vs placebo at end of Part 1. This yields an estimated long-term effect size of 0.99 ("large", defined as 0.8-1.19).

Table S19: FLACC: Pain

FLACC pain scale	**Baseline**	**End Part 1**	**End Part 2**	**End Part 3**
Group A	5.67	1.67	1.67	1.67
Group B	5.67	3.00	1.67	0.67
p-value*		0.53		
Cohen's d		0.56		
Estimated Cohen's d (Group A and B combined)			0.71	0.99

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.

Table 19-A: Percent Change in Average FLACC-Pain Compared to Baseline

	End of Part 1	**End of Part 2**	**End of Part 3**
Group A	-71%	-71%	-71%
Group B	-47%	-71%	-88%

4. Parent Global Impressions – Pitt Hopkins (PGI-PTHS): The PGI-PTHS evaluates 20 autism-related symptoms and 9 symptoms specific to individuals with PTHS. An initial severity score is calculated for each item on a scale of 0 (no symptom) to 6 (extremely severe). At end of Part 1, 2, and 3, a Likert scale is used to rate the change in severity compared to baseline, from much worse (- 3) to no change (0) to slightly better (+1), better (+2), and much better (+3).

4.1 PGI-PTHS: Average of 29 symptoms:
Baseline: The average severity of each symptom was calculated, including a score of "0" for patients who reported not having that symptom. In addition, a count was included to show

the number of patients with this symptom, Table S20. Children with PTHS had many significant symptoms at baseline. The largest impairments were in autism symptoms (language, sociability, stimming/perseveration), play skills, cognition, stools/GI, anxiety, gross motor, fine motor, and ataxia.

Patients who did not exhibit a particular symptom at baseline reported the severity as none "0". If these same patients reported "no change" (0) at subsequent time points, their follow-up score was reclassified as Not Applicable (NA) for analysis so that symptom severity change only included the patients who experienced that symptom at baseline.

Table S20: Severity of symptoms at Baseline on a Scale of 0 (No Symptom) to 6 (Extremely Severe) Sorted from highest to lowest severity for both group A and group B with Average and Overall severity at the bottom.

In general these children had a wide range of mental and physical symptoms, with 22 symptoms having a severity of 3 (moderate) or higher, and 8 symptoms having a severity of 5 (severe) or higher.

Symptom	Group A	Group B	Average of Group A and Group B	Baseline number of patients who have this symptom
Expressive Language	5.7	6.0	5.8	6
Stool/GI Problems	6.0	5.3	5.7	6
Fine Motor	5.0	5.7	5.3	6
Gross Motor	4.7	5.7	5.2	6
Ataxia	4.7	5.7	5.2	6
Play Skills	4.7	5.3	5.0	6
Cognition/thinking	6.0	3.7	4.8	6
Sociability	4.7	4.7	4.7	6
Anxiety	5.3	4.0	4.7	6
Stimming/perseveration	5.0	4.0	4.5	6
Attention/focus	4.7	4.0	4.3	6
Hyperactivity	4.7	4.0	4.3	6
Receptive Language	5.3	3.3	4.3	6
Sensory sensitivity	4.3	3.3	3.8	6
Sleep	3.7	3.7	3.7	6
Eye Contact	3.3	3.7	3.5	6
Irritability/mood	3.3	3.3	3.3	6
Tantrums/meltdowns	3.7	2.0	2.8	6
Aggression	3.3	1.7	2.5	5
Environmental allergies	2.3	2.7	2.5	6
Self-abuse	3.7	1.0	2.3	4
Hyperventilation	3.3	0.7	2.0	5
Flushing	2.0	1.7	1.8	6
Food allergies	0.7	3.0	1.8	3

Rashes	2.3	1.0	1.7	4
Self-limited diet	2.0	1.0	1.5	4
Daytime apnea	2.7	0.3	1.5	3
Seizures	1.3	1.3	1.3	2
Overall Symptoms	5.3	5.7	5.5	6
Average of all Symptoms	3.9	3.4	3.6	6

4.2 PGI-PTHS: Average of All Symptoms: The average of all symptoms represents changes in symptom severity only for patients who experienced that symptom. At the end of Part 1, Group A reported small improvements on the Average Score (average of all symptoms), more than 2x the improvement of Group B. The effect size was of medium size (Cohen's d=0.53) **Table S21.**

At the end of Part 2, both groups improved slightly more, and at the end of Part 3 there were slightly more improvements. The estimated effect size was medium of size (Estimated Cohen's d=0.52) **Table S21.**

By the end of Part 3, Group A averaged more than "slightly better" on the 27 symptoms, and Group B averaged about 60% of the way towards "slightly better" on the 27 symptoms. The estimated effect size is large (Estimated Cohen's d=1.03) **Table S21.**

Table S21: Change in the Average PGI-PTHS Score (Averaging Over 29 Symptoms Relative to Baseline

	End Part 1	**End Part 2**	**End Part 3**
Group A	0.70	0.88	1.27
Group B	0.32	0.45	0.69
p-value*	0.30		
Cohen's d	0.53		
Estimated Cohen's d (Group A and B combined)		0.52	1.03

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1

4.3 PGI-PTHS – Single Question on Overall:

This is a single question to assess the patient's overall symptom severity.

Part 1: Group A improved somewhat more than Group B. The effect size is small (Cohen's d=0.36) **Table S22.**

Part 2: Group A improved more, but Group B did not change. The estimated effect size was small (Estimated Cohen's d=0.70) **Table S22.**

Part 3: Both Groups improved slightly more. Group A averaged "better" and Group B averaged

"slightly better." The estimated Cohen's d was large (Estimated Cohen's d=0.97) **Table S22.**

These results are similar to the "average of 29 symptoms", but the magnitude of improvement is substantially more.

Table S22: PGI-PTHS Average Change in Overall Severity (Single Question)

	End Part 1	**End Part 2**	**End Part 3**
Group A	1.0	1.8	2.2
Group B	0.7	0.7	1.0
p-value*	0.36		
Cohen's d	0.47		
Estimated Cohen's d (Group A and B combined)		0.70	0.97

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.

4.4 PGI-PTHS- single question on stool/GI problems:

Due to the focus of this study on GI symptoms, we also report the single question on the PGI-PTHS about stool/GI problems.

At the end of Part 1, Group A improved approximately 2x as much as Group B. The effect size was small (Cohen's d= 0.49) Table S23.

At the end of Part 2, both Groups reported substantial improvement compared to baseline, and those improvements remained at the end of Part 3. So, Group A averaged approximately "better" and Group B

averaged between "slightly better" and "better." The estimated effect size for end of part 2 and end of part 3 was very large (Estimated Cohen's d=1.26 and 1.36, respectively) Table S21.

Table S23: Single Question on Stool/GI Problems

	End Part 1	**End Part 2**	**End Part 3**
Group A	1.3	1.8	2.2
Group B	0.7	2.0	1.7
p-value*	0.29		
Cohen's d	0.49		
Estimated Cohen's d (Group A and B combined)		1.26	1.36

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1

4.5 PGI-PTHS – individual symptoms: Although the largest improvement was in GI symptoms, parents reported almost as much improvement in several other symptoms, and some improvements on many other symptoms. Below is the average level of improvement for each

symptom on the PGI-PTHS, sorted from most to least improved, on a scale of + 3 (much better) to -3 (much worse). No average scores were below zero.

Table S24: Average Improvement at End of Part 3

Symptom	End of Part 3 Average Improvement	End of Part 3 (Patients who have this symptom)
Rashes	2.0	4
Stool/GI Problems	1.9	6
Irritability/mood	1.8	6
Seizures	1.5	2
Sociability	1.4	6
Anxiety	1.4	6
Tantrums/meltdowns	1.3	6
Aggression	1.2	5
Gross Motor	1.2	6
Stimming/perseveration	1.1	6
Receptive Language	1.0	6
Eye Contact	1.0	6
Fine Motor	0.9	6
Daytime apnea	0.9	3
Expressive Language	0.8	6
Self-abuse	0.8	4
Attention/focus	0.8	6
Hyperactivity	0.8	6
Ataxia	0.8	6
Cognition/thinking	0.7	6
Sensory sensitivity	0.7	6
Hyperventilation	0.6	5
Sleep	0.5	6
Food allergies	0.5	3
Flushing	0.5	6
Play Skills	0.3	6
Environmental allergies	0.3	6
Self-limited diet	0.3	4
Overall Symptoms	**1.6**	**6**
Average of all symptoms	**1.0**	**6**

Interpretation: Improvements in GI symptoms and pain likely explain some of the improvements in PTHS symptoms, including the improvements in irritability, sociability, anxiety, tantrums, and aggression. We hypothesize some improvements may also be due to a reduction of neurotoxins from pathogenic bacteria.

Table S24: PGI-PTHS: End of Part 3 (Groups A and B) vs Placebo (Group B at End of Part 1) – Decrease in PGI-PTHS severity score for each of 29 symptoms and an

Average of all symptoms

Average Decrease In Severity - PGI-PTHS	End of Part 3 (Group A and Group B)	Placebo End of Part 1	Number of patients in Group A and B with that symptom at baseline	Number of patients in Group B with that symptom at baseline
Expressive Language	0.8	0.0	6	3
Receptive Language	1.0	0.7	6	3
Play Skills	0.3	0.0	6	3
Cognition/thinking	0.7	0.3	6	3
Attention/focus	0.8	0.7	6	3
Stool/GI Problems	1.9	0.7	6	3
Sleep	0.5	0.3	6	3
Sociability	1.4	0.7	6	3
Eye Contact	1.0	0.3	6	3
Hyperactivity	0.8	0.7	6	3
Tantrums/meltdowns	1.3	0.0	6	3
Irritability/mood	1.8	1.0	6	3
Anxiety	1.4	0.7	6	3
Stimming/perseveration	1.1	0.0	6	3
Sensory sensitivity	0.7	0.0	6	3
Aggression	1.2	0.5	5	2
Self-abuse	0.8	0.0	4	1
Seizures	1.5	0.0	2	1
Self-limited diet	0.3	0.0	4	1
Hyperventilation	0.6	0.0	5	2
Daytime apnea	0.9	0.0	3	1
Gross Motor	1.2	0.7	6	3
Fine Motor	0.9	0.3	6	3
Ataxia	0.8	0.3	6	3
Food allergies	0.5	0.0	3	2
Environmental allergies	0.3	0.0	6	3
Flushing	0.5	0.0	6	3
Rashes	2.0	0.0	4	1
Overall Symptoms	1.6	0.7	6	3
Average of all symptoms	1.0	0.3	6	3

Table S25: Effect Size at End of Part 1 (Group A vs Group B) and estimated effect size at End of Part 3 (Group A and B vs Group B at End of Part 1)

Cohen's d

SYMPTOM	End of Part 1: Effect Size	End of Part 3: Estimated Effect Size	End of Part 1	End of Part 3
Daytime Apnea (short periods without breathing)	2.12	1.12	Group A = 2 Group B = 1	Group A = 2 Group B = 1
Expressive Language/Speech	1.63	0.74	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Hyperventilation	1.22	0.80	Group A = 3 Group B = 2	Group A = 3 Group B = 2
Eye Contact	0.82	0.60	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Gross Motor Skills	0.82	0.79	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Play Skills	0.82	0.48	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Tantrums/Meltdowns	0.82	1.18	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Stimming/Perseveration	0.82	1.00	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Sensory Sensitivity	0.82	0.97	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Environmental Allergies	0.82	0.48	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Self-Abuse	0.58	0.83	Group A = 3 Group B = 1	Group A = 3 Group B = 1
Average of all symptoms	0.53	1.03	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Stools/GI Problems	0.49	1.36	Group A = 3 Group B = 3	Group A = 3 Group B = 3
OVERALL Symptoms	0.47	0.97	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Attention/Focus	0.41	0.24	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Fine Motor Skills	0.41	0.57	Group A = 3 Group B = 3	Group A = 3 Group B = 3

Ataxia (lack of communication between brain and body)	0.41	0.56	Group A = 3 Group B = 3	Group A = 3 Group B = 3
			3	3
Cognition/Thinking	0.37	0.44	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Irritability/Mood	0.31	0.71	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Sociability	0.26	0.67	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Sleep	0.25	0.15	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Rashes	0.22	1.16	Group A = 3 Group B = 1	Group A = 3 Group B = 1
Aggression	0.16	0.78	Group A = 3 Group B = 2	Group A = 3 Group B = 2
Receptive Language	0.00	0.34	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Hyperactivity	0.00	0.16	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Anxiety	0.00	0.67	Group A = 3 Group B = 3	Group A = 3 Group B = 3
Seizures	Unable to calculate	Unable to calculate	Group A = 1 Group B = 1	Group A = 1 Group B = 1
Food Allergies	No change	0.71	Group A = 1 Group B = 2	Group A = 1 Group B = 2
Self-limited Diet	No change	0.50	Group A = 3 Group B = 1	Group A = 3 Group B = 1
Flushing	No Change	0.48	Group A = 3 Group B = 3	Group A = 3 Group B = 3

5. Clinical Global Impressions

5.1 Clinical Global Impressions GI (CGI-GI):

The CGI-GI was rated by the study physician, including Severity Scale, Improvement Scale, and Efficacy Index. It was rated at Baseline, end of Part 1, and end of Part 2.

Severity Scale: The severity is rated on a 7-point scale:

1. Normal, not at all ill

2. Borderline mentally ill
3. Mildly ill
4. Moderately ill
5. Markedly ill
6. Severely ill
7. Among the most extremely ill patients

In Part 1, both Groups improved slightly **Table S26**.

In Part 2, both Groups improved more. However, some GI symptoms remained **Table S26**.

The calculated effect size (Cohen's d) at the end of Part 1 was 0.31 (where a "small" effect size is defined as 0.20 to 0.49) **Table S26**.

Table S26: CGI-GI - Severity

CGI-GI Severity	**Baseline**	**End Part 1**	**End Part 2**
Group A	6.7	5.7	5.0
Group B	6.0	5.3	4.0
p-value*		0.36	
Cohen's d		0.31	

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.

Improvement Scale: The improvement scale uses the following definitions

1. Very much improved
2. Much improved
3. Minimally improved
4. No change
5. Minimally worse
6. Much worse
7. Very much worse

In Part 1, the treatment Group improved slightly more than placebo Table S27.

In Part 2, both Groups improved more, with greater improvement for Group B after they were treated Table S27

The calculated effect size (Cohen's d) at the end of Part 1 was 0.31 (small) Table S25.

Table S27: CGI-GI-Improvement: note that a score of 4 is no change, and that lower scores indicate more improvement.

CGI-GI - Improvement	**End Part 1**	**End Part 2**
Group A	3.0	2.7
Group B	3.3	2.0

p-value*	0.36	
Cohen's D	0.31	

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.

Efficacy Index

The efficacy index is a grid plotting therapeutic effect vs. side effects.

The study physician did not note any significant side effects during Part 1 or Part 2 of the treatment, and all patients were marked as "none" at both timepoints.

Table 28: CGI-GI Side Effect

CGI-GI-side effect	End Part 1	End Part 2
Group A		
PH-006	None	None
PH-009	None	None
PH-003	None	None
Group B		
PH-002	None	None
PH-004	None	None
PH-005	None	None

Therefore, for ease of analysis, we plot only the average "Therapeutic Effect", defined as:
Marked: 3
Moderate: 2
Minimal: 1
Unchanged or worse: 0

In Part 1, the treatment group improved more than the placebo group **Table S29**.

In Part 2, Group A remained improved, and the Group B improved more Table S29. The calculated effect size (Cohen's d) at the end of Part 1 was 0.31 (small) Table S29.

Table S29: CGI-GI-Therapeutic Effect

CGI-GI Therapeutic Effect	End Part 1	End Part 2
Group A	1.0	1.3
Group B	0.7	2.0
p-value*	0.36	
Cohen's d	0.31	

5.2 Clinical Global Impressions – PTHS:

The CGI-PTHS was rated by the study physician, including Severity Scale, Improvement Scale, and Efficacy Index.
Severity Scale: The severity is rated on a 7-point scale:

1) Normal, not at all ill
2) Borderline mentally ill
3) Mildly ill
4) Moderately ill
5) Markedly ill
6) Severely ill
7) Among the most extremely ill patients

In part 1, Group A improved slightly more than Group B Table S30.

In part 2, both Groups improved more Table S30.

The calculated effect size (Cohen's d) at the end of Part 1 was 0.37 (small) Table S30. CGI-PTHS – Severity

Table S30: CGI-PTHS – Severity

CGI-PTHS-Severity	**Baseline**	**End Part 1**	**End Part 2**
Group A	6.3	5.7	4.7
Group B	5.7	5.3	4.7
p-value*		0.34	
Cohen's d		0.37	

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.

<u>Improvement Scale</u>

The improvement scale uses the following definitions

1. Very much improved
2. Much improved
3. Minimally improved
4. No change
5. Minimally worse
6. Much worse
7. Very much worse

In Part 1, Group A improved more than Group B, Table S31.

In Part 2, Group A improved more, and Group B improved slightly more, Table S31 The calculated effect size (Cohen's d) at the end of Part 1 was 0.82 (large), Table S31.

Table S31: CGI-PTHS-Improvement

CGI-PTHS - Improvement	End Part 1	End Part 2
Group A	3.0	2.3
Group B	3.7	3.3
p-value*	0.19	
Cohen's d	0.82	

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.

Efficacy Index

The efficacy index is a grid plotting therapeutic effect vs. side effects.

The study physician did not note any significant side effects during Part 1 or Part 2 of the treatment, and all patients were marked as "none" at both timepoints.

Table S32: CGI-PTHS - Side Effect

CGI-PTHS-side effect	End Part 1	End Part 2
Group A		
PH-006	None	None
PH-009	None	None
PH-003	None	None
Group B		
PH-002	None	None
PH-004	None	None
PH-005	None	None

Therefore, for ease of analysis, we plot only the average "Therapeutic Effect",
defined as:
Marked: 3
Moderate: 2
Minimal: 1
Unchanged or worse: 0

In Part 1, Group A had much more improvement than Group B Table S33.

In part 2, Group A remained improved, and Group B had some small improvements Table S33. The calculated effect size (Cohen's d) at the end of Part 1 was 0.87 (large) Table S33.

Table S33: CGI-PTHS-Therapeutic Effect

CGI-PTHS Therapeutic Effect	End Part 1	End Part 2
Group A	1.3	1.3
Group B	0.3	0.7
p-value*	0.19	
Cohen's d	0.87	

*T-Test (1-sided) comparison of the change in Group A vs. the change in Group B at the end of Part 1.

6. Occupational Therapy:

The Sensory Processing Measure (SPM), developed by Western Psychological Services (WPS), is a widely used tool by occupational therapists to assess sensory integration challenges. At baseline, an occupational therapist used the SPM as part of a comprehensive physical, sensory, and functional evaluation.

Questions are rated on a Likert scale based on frequency (never, occasionally, frequently, and always). Raw scores are converted to standard scores and interpreted in terms of typical performance (40T-59T), some problem (60T-69T) or definite dysfunction (70T-80T).

The occupational therapist was not able to continue assisting in the study due to COVID-19. In addition, we do not have baseline data for two out of three patients from Group B, due to loss of data at the start of COVID.

At baseline, Group A had a lower score than Group B.

At end of part 1, Group A showed slight worsening (2.5%) while Group B showed slight improvement (2.4%).

At the end of part 2, Group A showed slightly more worsening while Group B improved slightly.

At end of part 3, Group A was slightly worse compared to baseline (3.3% worse) while Group B showed slight improvement (3.8%).

Table S34: OT Evaluation – Average Total Score of SPM

Average Total Score (Does not Include SOC and PLA)	Baseline	End of Part 1	End of Part 2	End of Part 3
Group A	66.3	68	70	68.5
Group B	71	69.3	68	68.3
P-Value*		-		
Cohen's d*		-		

*P-value and Cohen's -d cannot be calculated since Group B only had data for 1 patient.

References:

Malviya S, Voepel-Lewis T, Burke C, Merkel S, Tait AR. The revised FLACC observational pain tool: improved reliability and validity for pain assessment in children with cognitive impairment. Paediatr Anaesth. 2006 Mar;16(3):258-65. doi: 10.1111/j.1460-9592.2005.01773.x. PMID: 16490089.