STATISTICAL ANALYSIS PLAN

NeurOmodulaTion for Accidental Bowel LEakage (NOTABLE)

Short title: NOTABLe
ClinicalTrials.gov NCT Number: NCT03278613

SAP VERSION: Version 1.0

SAP DATE: February 28, 2020

SPONSOR: NICHD

PREPARED BY: RTI International.
3040 Cornwallis Rd
Research Triangle Park, NC 27709-2104

AUTHOR (S): Benjamin Carper
Marie Gantz

Contents

LIST OF ABBREVIATIONS

ABC	Anticholinergic versus Botox Comparison trial
ATLAS	Ambulatory Treatments for Leakage Associated with Stress Incontinence trial
BBUSQ	Birmingham Bowel Urinary Symptom Questionnaire
BD	Bladder diary
BE- DRI	Behavior Enhances Drug Reduction of Incontinence trial
BPTx	Behavioral/pelvic floor therapy
CDF	Cumulative distribution function
CST	Cough stress test
DCC	Data Coordinating Center
DO	Detrusor overactivity
DSMB	Data and Safety Monitoring Board
EQ-5D	European Quality of Life-5 Dimensions
HRQOL	Health related quality of life
IE	Incontinence episode
ICI	International Consultation on Incontinence
ICS	International Continence Society
IIQ	Incontinence Impact Questionnaire
IRB	Institutional Review Board
ITT	Intention-to-treat
IUGA	International Urogynecological Association
MESA	Medical, Epidemiologic, and Social Aspects of Aging
MID	Minimum important difference
MIMOSA	Mixed Incontinence: Medical or Surgical Approach trial
MSM	Medical Safety Monitor
MUI	Mixed urinary incontinence
MUS	Mid-urethral sling
OAB	Overactive bladder
OAB-q	Overactive Bladder Questionnaire
OAB-q-SS	Overactive Bladder Questionnaire-Symptom subscale
OAB-SAT-q	Overactive Bladder Questionnaire-Satisfaction with Treatment Questionnaire
OPTIMAL	Operations and Pelvic Muscle Training in the Management of Apical Support Loss trial
PFD	Pelvic floor disorder
PFDI	Pelvic Floor Disorder Inventory
PFDN	Pelvic Floor Disorders Network
PFMT	Pelvic floor muscle training
PGI-I	Patient Global Impression- Improvement
PGI-S	Patient Global Impression-Severity

PISQ	Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire
POPQ	Pelvic Organ Prolapse Quantification system
PVR	Postvoid residual
QoL	Quality of life
QUID	Questionnaire for Urinary Incontinence Diagnosis
RCT	Randomized controlled trial
RUBI	Refractory idiopathic urge incontinence and botulinum A injection trial
SAE	Serious adverse event
SD	Standard deviation
SISTEr	Stress Incontinence Surgical Treatment Efficacy Trial
SUI	Stress urinary incontinence
TOMUS	Trial of Mid-Urethral Slings
TOT	Transobturator tape sling
TVT	Tension-free vaginal tape sling
TVT-O	Tension-free vaginal tape obturator
UDE	Urodynamic evaluation
UDI	Urogenital Distress Inventory
UI	Urinary incontinence
UIE	Urinary incontinence episode
UITN	Urinary Incontinence Treatment Network
UUI	Urge urinary incontinence
ValUE	Value of Urodynamic Evaluation trial
VPFMC	Voluntary pelvic floor muscle contraction
3IQ	3 Incontinence Questions Assessment Tool

1 BACKGROUND AND PROTOCOL HISTORY

Accidental bowel leakage (ABL), aka fecal or anal incontinence, is a common debilitating condition experienced by millions of women. Unfortunately, there has been little emphasis on identification and treatment of this condition. With new findings confirming the high prevalence of the disorder and new treatment options available, there is a move toward educating the public about this under-recognized problem. The overarching goal of this randomized trial is to determine if percutaneous tibial nerve stimulation (PTNS) is more effective than sham PTNS for the treatment of fecal incontinence (FI) in women. Secondary objectives include assessing the durability of effect of the PTNS treatment, assess the impact of completing a bowel diary and receiving education on the severity of symptoms, and assess the ability of the PFDN ABL phone app to detect changes from baseline in fecal incontinence episode.

2 PURPOSE OF THE ANALYSES

This statistical analysis plan (SAP) contains detailed information about statistical analyses to be performed to address the primary and secondary aims of NOTABLE. All analyses that will be included in the primary manuscript are described. Additional exploratory analyses may be performed to support further manuscript development. These analyses will not require an update to the SAP.

3 STUDY AIMS AND OUTCOMES

3.1 Study Aims

The overarching goal of this randomized clinical trial is to determine if PTNS is more effective than sham PTNS for the treatment of FI in women after 12 weeks of treatment and 1 year after initiating treatment.

3.1.1 Primary Aims

The primary aim of this study is to determine whether the change from baseline in St. Mark's (Vaizey) score in women with symptomatic ABL undergoing PTNS differs from sham, after 12 weeks of stimulation.

3.1.2 Secondary Aims

Secondary aims of this study include the following:

1. To compare changes from baseline in self-reported functional outcomes after 12 weekly stimulation sessions in both PTNS and sham groups. Outcomes include:
   - Measures of symptom severity documented on a 14-day bowel diary including ABL episodes, fecal urgency, and number of defecation(s)/day, number of FIE-free days per week.
   - Condition-specific quality of life (QOL); global impression of improvement; co-existent bowel, bladder, prolapse symptoms; adaptation measures; sexual function; and change in ongoing interventions for ABL described at baseline.
2. Durability of Effect (Part II): To determine whether symptom relief amongst study "responders" can be sustained for one year with maintenance treatments. A 4-point reduction from baseline in the St. Mark's score will be considered clinically significant and will be used to define an eligible treatment response for entry into Part II.

   a. To determine whether either a fixed schedule or a subject driven (PRN) schedule of treatments in "responders" of Part I is effective in maintaining symptom control at one year from start of treatments.
      This aim will estimate the percent of women assigned to each treatment schedule who maintain symptom control (defined as a 4-point reduction from baseline in the St. Mark's score) at one year from start of PTNS treatments.
   b. To determine whether the durability of symptom reduction acquired in Part I is sustained for up to 9 months after starting a maintenance strategy (comparison of Part I and Part II outcomes).
   c. To determine if the Fixed and the PRN schedule of treatments in Part II are feasible to implement, and are associated with different costs and participant satisfaction.
3. 6-Month Follow-Up of Symptom Control, Improvement, and QOL After Final PTNS Session Amendment to Protocol
   a. To establish a 6-month follow-up phase after the final PTNS session to determine the duration of effect of treatment. Follow-up will be discontinued when subjects report PGSC of ≤2 or report initiating NEW treatment for ABL that is prescribed by a health provider.
4. Impact of completing a bowel diary and receiving education on fecal incontinence (FI) (NIDDK pamphlet) on symptom severity:
   a. To determine the impact of education and completing a bowel diary on FI symptom severity as measured by a change from baseline in the St. Mark's score and change in weekly FIE (Week 1 vs. Week 4) during the Run-In Phase.
   b. To compare modalities of the bowel diary (phone app vs. paper) on their impact on FI symptom severity as measured by change from the baseline in the St. Mark's score and change in mean FIEs/week recorded in Week 1 and Week 4 of the Run-In Phase.
5. Ability of the PFDN ABL phone app diary to detect change: To determine if the changes from baseline in FIEs recorded on the PFDN ABL phone app diary correlate with changes from baseline in other measures of FI symptom severity including the St. Mark's score after 12 weeks of stimulation.
6. Determine association between St. Mark's score and both fat and fiber intake alone and in combination
   a. The overall goal is to determine if accidental bowel leakage severity is associated with dietary intake. Dietary intake will be determined using the Dietary Screener self-administered questionnaire which captures both dietary fats and fiber (68). The association between St. Mark's score and lower dietary fat intake and higher fiber intake will be assessed.

3.1.3 Exploratory Aims

Exploratory aims of this study include the following:

1. Safety: To describe and compare adverse events in PTNS and sham groups.
2. Treatment Compliance:
   a. To describe adherence to the treatment protocol in the PTNS group of Part I (defined as attendance at 10 of 12 scheduled treatment sessions) and to compare it to that of the sham group.
   b. To describe treatment adherence in the fixed schedule group of Part II.

3. Willingness to continue maintenance PTNS therapy: Amongst responders, to determine their willingness to continue PTNS therapy for maintenance of symptom suppression in Part II.
4. Validity of the sham: To determine whether participants were aware of their assigned intervention in Part I.
5. Predictors of response: To identify clinical characteristics associated with treatment success at the end of Part I and Part II. Characteristics to be studied include, but are not limited to, age, BMI, Baseline ABL (severity), Baseline St. Mark's (Vaizey) score, Baseline stool consistency, and adherence to treatment schedule.
6. 6. Rate of UTIs: To identify if there is a difference in the rate of UTIs treated with antibiotics between the PTNS and sham groups.

3.2 Outcomes

3.2.1 Primary Outcomes

The primary outcomes for this study are the following:

a. Part I. Change from baseline in St. Mark's (Vaizey) score after 12 weeks of stimulation.
b. Part II. Percent of "responders" at one year.

3.2.2 Secondary and Exploratory Outcomes

Secondary outcome measures were selected with a priority to minimize redundancy within the instruments and subject burden. Additionally, the selected measures will enable comparison of study findings to other published studies of PTNS for FI. They include measures of symptom severity, QOL, common co-existent bowel and bladder symptoms, global impression of improvement, and behavior adaptations for pelvic floor disorders.

Secondary/ Exploratory Aims	Measure	Outcome
PTNS and sham group differences through 13 weeks in		
Self-reported Symptom Severity, Bother, and Functional Outcomes	14-day bowel diary	Change from baseline in: episodes of FI, fecal urgency, defecation
	ABLE	Change from baseline in score
Condition-specific QOL	FIQL, Modified Manchester Questionnaire with FSFI	Change from baseline in scores
Other Bowel, Bladder, and Prolapse Symptoms	PFDI-20, PFIQ-7, PAC-SYM	Change from baseline in scores
Sexual Function	PISQ-IR	Change from baseline in score
Adaptive Behaviors	Fecal Incontinence Adaptation Index	Change from baseline in score
Use of Supplementary Treatments for ABL	St. Mark's (Vaizey) question 3b	Self-reported use of constipating Rx over time
Patient Global Impression of Improvement (PGI-I) and Patient Global Symptom Control (PGSC)	PGI-I modified for bowel function, PGSC Rating (modified for ABL)	PGI-I and PGSC scores over time
General Health Survey	SF-12	Change from baseline
Safety/Adverse Events	SAE/AE reports	AEs and SAEs reported
Other Secondary/Exploratory Measures		

PTNS Treatment Adherence and Feasibility of Treatment Schedule	Session attendance. Session completion (30 min stimulation)	% who attend 10 of 12 scheduled PTNS treatments in Part I and 9 of 11 sessions in Part II (Fixed schedule)
Validity of the Sham (Part I)	Query on treatment assignment	% correctly identifying their treatment assignment
Effectiveness of PTNS Maintenance Schedules	St. Mark's (Vaizey) score and other functional outcome measures in Part I	Change from baseline in St. Mark's (Vaizey) scores through 1 year; changes in other functional outcome measures through 1 year
Rate of UTIs treated with antibiotics	Medical Follow-Up Form	Difference in rate between the PTNS and Sham Groups
Cost of PTNS Maintenance Schedules	Treatment Form	Mean # of PTNS treatments/year in Part II Mean Treatment interval for those receiving PTNS in Part II

3.2.3 Safety Outcomes

Safety outcomes will be assessed in a descriptive manner at each DSMB meeting without formal statistical tests.

4 STUDY METHODS

4.1 Overall Study Design and Plan

The study is a multi-center, randomized, clinical trial of women with FI. This is a three-part trial with a Run-In Phase prior to randomization in Part I.

The purpose of the Run-In Phase is to identify and exclude from randomization women whose FI symptoms improve to the point that they are below the eligibility threshold of 12 points on the St. Mark's scale after receiving education on FI and completing two 7-day bowel diaries. The Run-In Phase is a 4-week period during which participants will complete bowel diaries (either on paper or by phone app) in Weeks 1 and 4. Women who have a St. Mark's score ≥ 12 after the Run-In will move on to Part I of the study.

Part I is a randomized, single-masked controlled trial comparing the effectiveness of PTNS to a validated sham in women with refractory FI using a 2:1 assignment. Its purpose is to evaluate the impact of PTNS above sham for treatment of ABL symptoms. The primary outcome, change from baseline in the St. Mark's (Vaizey) score, will be measured over time (after 12 weeks of stimulation). A change of 4 points will be considered clinically significant. Randomization will be by randomly permuted blocks, stratified by site and by type of diary completed in the Run-In Phase (paper or phone app). The analysis of Part I outcomes will occur when the Part I sample size is reached and those participants have completed 12 weeks of stimulation.

Part II is a trial of two maintenance strategies amongst participants who are treatment responders, where treatment response is defined as an improvement from baseline in St. Mark's score of ≥ 4 points after 12 weeks of stimulation. Eligible participants will be randomized using a 1:1

assignment to either a fixed schedule of treatments or a patient symptom driven (PRN) treatment schedule. The purpose of Part II is to evaluate:

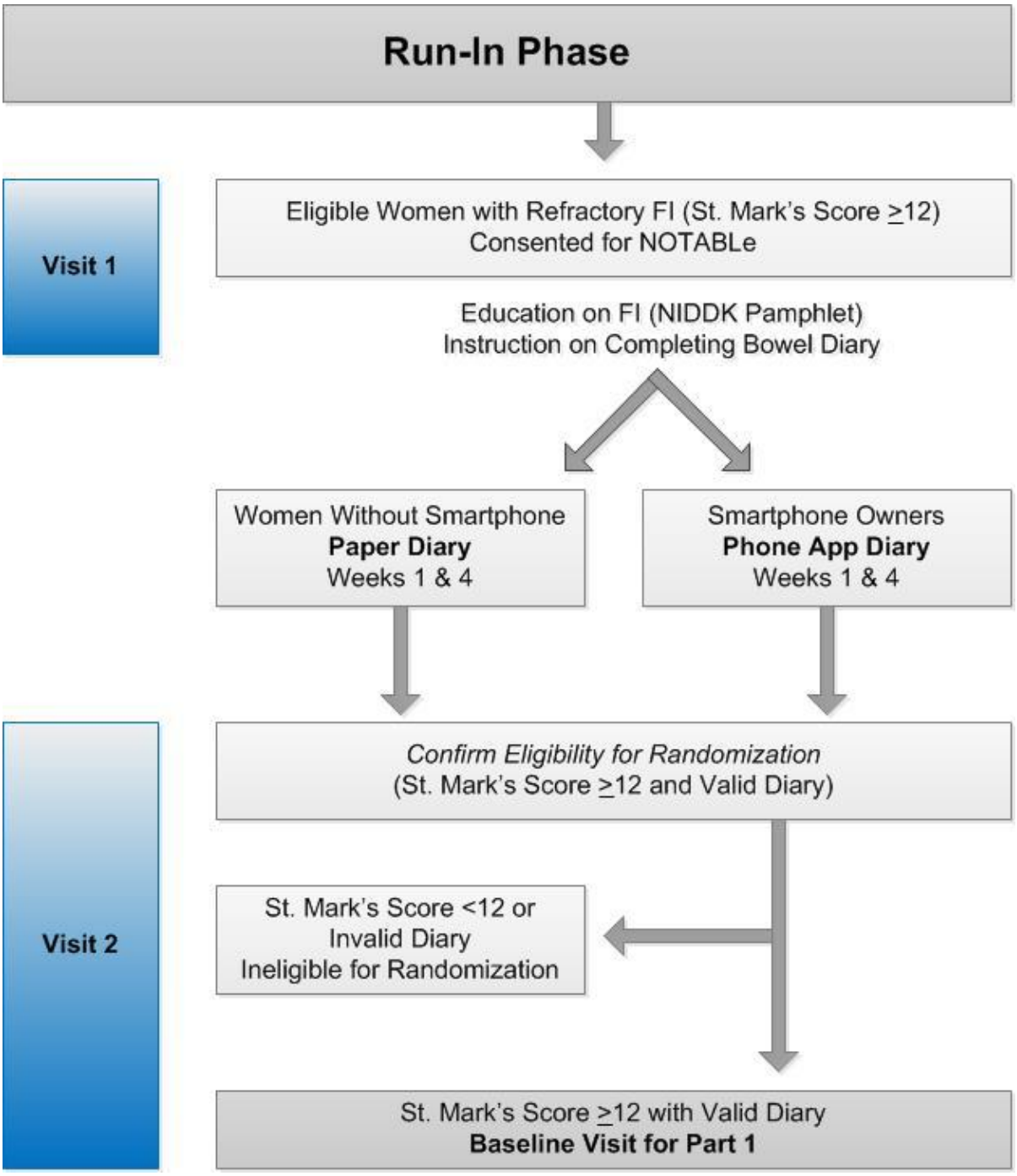
a. The percent of women assigned to each treatment schedule who maintain symptom control (defined as a 4-point reduction from baseline in the St. Mark's score) at one year from start of treatment.
b. Whether the fixed and the PRN schedule of treatments in Part II are feasible to implement, and are associated with different costs and/or participant satisfaction

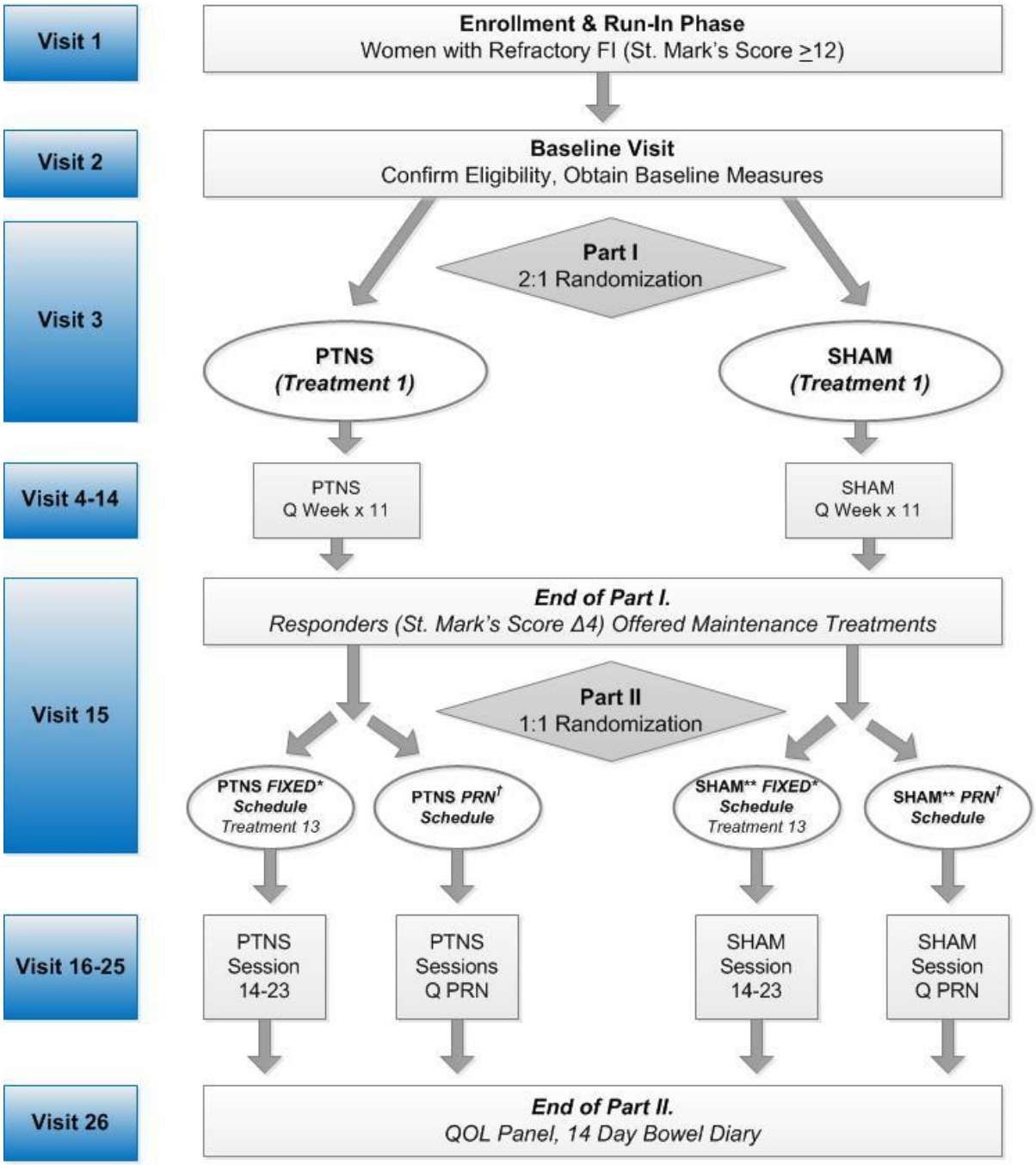
Randomization to Part II maintenance strategy will be by randomly permuted blocks. In order to maintain masking prior to the analysis of Part I data, Part I participants who respond to either PTNS or sham treatment will be eligible to begin Part II. Participants responding to sham in Part I will continue to receive sham treatments in Part II. Part II randomization will not be stratified by site but will be stratified by PTNS/sham group to assure that randomization of the PTNS group is balanced between the two maintenance groups. Only women receiving PTNS treatment will be included in Part II analyses.

The extension will follow women for up to an additional 6-months after the final PTNS session. In the absence of published data to indicate the duration of PTNS effect, the Protocol Committee proposes up to a 6-month observation period at 4-week intervals after visit 26: 6 weeks, 10 weeks, 14 weeks, 18 weeks, 22 weeks and 26 weeks after final PTNS session. Currently, the PRN group provides a PGSC score to the coordinators by phone in lieu of attending study visits. If their score does not trigger an in-person treatment visit, the subjects are emailed a secure online link to the NOTABLe REDCap database where they enter responses to the ePRO panel. This ePRO mechanism will be used in the follow-up phase to collect the abbreviated panel of outcome measures. All NOTABLe subjects randomized to Part II and with total of ≥6 months of PTNS treatment (≥6 months from randomization into Part I) will be eligible for inclusion. Any NOTABLe subjects with total of <6 months of PTNS sessions (those exiting Part II early) will be excluded from the 6-month extended follow-up. Subjects will be withdrawn from follow-up when they report PGSC of ≤2 or initiate NEW treatment for ABL that is prescribed by a health provider such as pessary (for ABL), Diphenoxylate/Atropine (Lomotil), sacral neuromodulation, or anal/colonic surgery. Resumption of baseline compensatory strategies such as Imodium, fiber supplements, pelvic muscle strengthening exercises are not considered NEW treatment for ABL during this observation period.

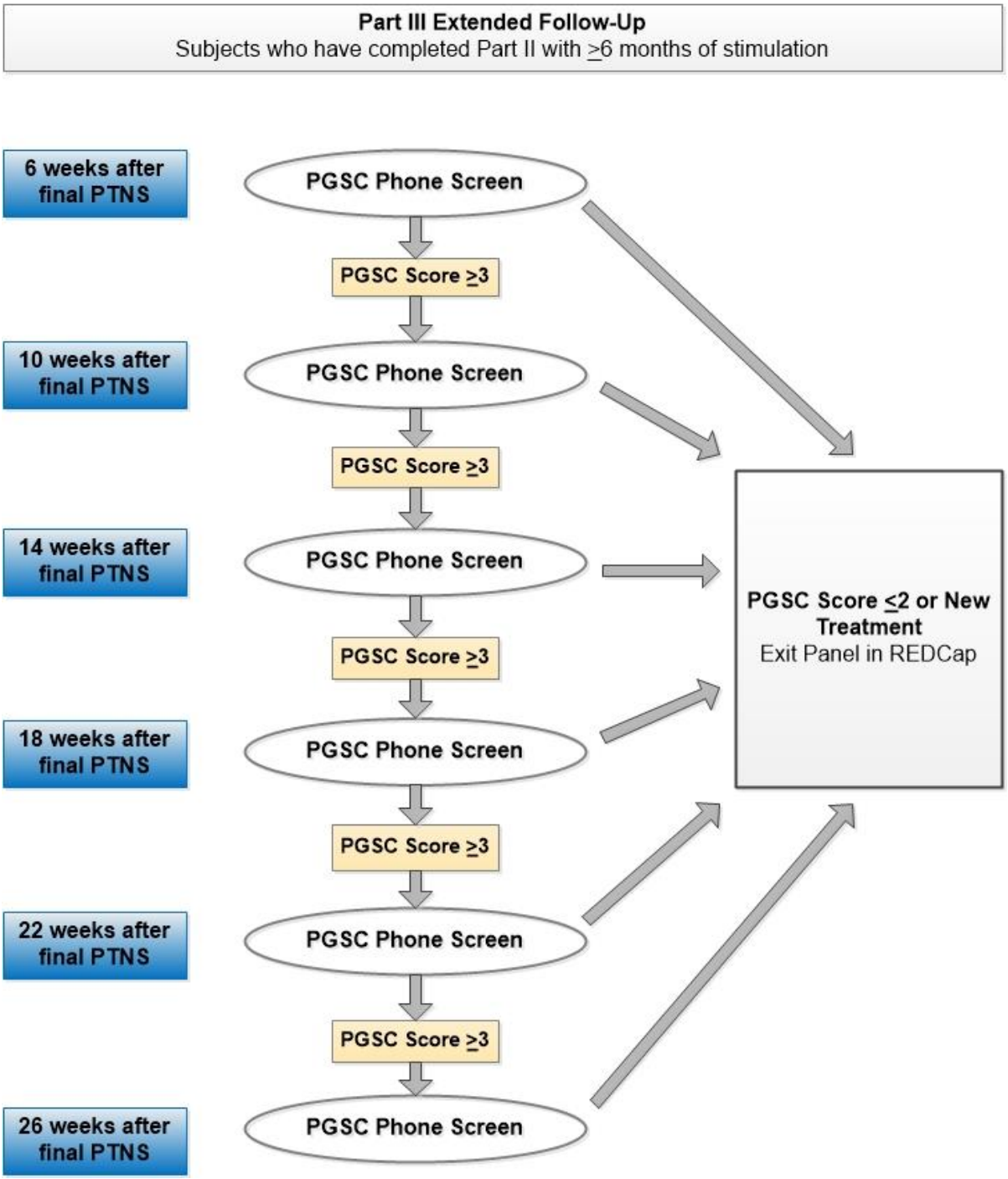
If Part I analysis does not demonstrate that PTNS is superior to sham for the treatment of FI, then the NOTABLe study will be halted and participants involved in Part II will be notified that the study is ending and maintenance treatments will not continue. If the analysis of Part I data supports the use of PTNS for FI, treatment of sham participants in Part II will be discontinued and additional participants will be recruited in order to reach the Part II sample size. These participants will complete the Run-In Phase and, if eligible after the Run-In Phase, they will receive PTNS treatments as specified in Part I. Women who are PTNS responders after 12 weeks of stimulation will be eligible for randomization to Part II. Women who previously completed Part I of the study in the sham group may be re-enrolled in this part of the study if they meet eligibility criteria. These women will be randomized in the sham group stratum of Part II, but their data will be included in Part II analysis as PTNS responders.

A figure illustrating the study design inclusive of the Run-In Phase is depicted on the next two pages.





* FIXED Schedule: Q 14d x 2; Q 21d x2; Q 28d x 7 (11 sessions in Phase II)
† PRN Schedule projected to be Q 1.1-1.3 per month (up to 12 sessions in Phase II)
**Sham sessions will be discontinued if PTNS demonstrates superiority to Sham



4.2 Study Population

Subjects will be women ≥ 18 years, with ≥ 3 months of refractory symptoms of ABL. In this study, subjects will have failed to achieve satisfactory symptom control from two first-line treatments for ABL: supervised pelvic muscle training (PMT) and constipating medications. The PMT must have been conducted by a health provider experienced with treating ABL (minimum 2 treatment sessions). The lack of response to constipating agents may be due to ineffectiveness,

intolerance, unwillingness, or contraindication. Supplementary fiber bulking will not be considered a treatment for ABL due to lack of data supporting its effectiveness for ABL.

4.2.1 Subject Characteristics

Inclusion and exclusion criteria for the NOTABLE trial are as follows:

Inclusion Criteria:
Participants must meet all the listed inclusion criteria by patient self-report.

1. Women ≥ 18 years of age
2. FI symptoms ≥ 3 months
3. Baseline St. Mark's score of ≥ 12
4. Attended ≥ 2 supervised PMT for ABL
5. Intolerance, unwillingness, or inadequate response to constipating medications
6. Current negative colon cancer screening based on the USPSTF's recommendation for colorectal cancer screening (2016)

Exclusion Criteria:

Subjects who meet any of the following criteria by patient self-report (or pregnancy test) are ineligible for enrollment in the study. Criteria pertaining to the site of the PTNS or sham needles or surface electrodes only exclude a participant from the study if she does not have an unaffected leg to which PTNS or sham can be applied.

1. Previous PTNS treatment
2. History of uncontrolled diarrhea in the past 3 months (usual or most common stool type over the preceding 3 months of 7 on the Bristol Stool Form Scale)
3. History of severe constipation in the past 3 months (1 on the Bristol Stool Form Scale)
4. History of inflammatory bowel disease (includes Crohn's disease and ulcerative colitis, but does not include irritable bowel disease)
5. Unrepaired rectovaginal fistula/chronic 4th degree laceration
6. Full thickness rectal prolapse
7. History of congenital anorectal malformation
8. History of bowel resection surgery for any indication
9. Minor anal procedures within 6 months for treatment of ABL (injection of bulking agent or radiofrequency energy) or ligation of hemorrhoids
10. Prior pelvic or abdominal radiation
11. Diagnosis of cancer of the descending colon or anus
12. Diagnosis of cancer in the region where the PTNS or sham needles or surface electrodes would be placed
13. Pacemaker, implantable defibrillator
14. Current use of Interstim sacral nerve stimulator or TENS in the pelvic region, back, or legs
15. Clinically significant neurological disorders known to affect anal continence
16. Coagulopathy
17. Severe peripheral edema preventing accurate placement of PTNS needles
18. Chronic swollen, infected, inflamed skin or skin eruptions (e.g., phlebitis, thrombophlebitis, varicose veins) in the region where the PTNS or sham needles or surface electrodes would be placed
19. Metal implant in foot/toes near TENS electrode location

20. Marked sensory deficit (numbness) of feet or ankles in the region where the PTNS or sham needles or surface electrodes would be placed
21. Childbirth within the last 3 months
22. Pregnant or planning to become pregnant during the study duration 1 year; a urine pregnancy test will be performed and must be negative by the first intervention visit if the participant is of childbearing potential
23. Unwilling to use acceptable form of contraceptive if the participant is of childbearing potential
24. Participation in another intervention trial impacting bowel function
25. Inability to provide informed consent, complete questionnaires independently, or to attend intervention sessions
26. Unable or unwilling to complete the bowel diary in Run-In Phase (valid diary defined as data from ≥ 10 of 14 days with minimum of 3 consecutive days per week)
27. Unwilling to download bowel diary app onto smartphone if the participant owns a smartphone
28. Visual impairment prohibiting reading the paper diary, the smart phone screen
29. Unable to speak, read, or write in English or Spanish at a basic level

To be eligible for randomization into Part I, enrolled subjects will have completed the 4-week Run-In Phase which includes receiving standardized verbal and written education on FI as delineated in the NIDDK brochure and have demonstrated their ability to complete a bowel diary (providing ≥ 10 of 14 days of diary data with minimum of 3 consecutive days per week). All subjects randomized in Part I will have a St. Mark's score of ≥ 12 points.

To be eligible for the 6-month extended follow-up after the final PTNS session, NOTABLe subjects will have been randomized to Part II and have completed a total of ≥6 months of PTNS treatment (≥6 months from randomization into Part I). NOTABLe subjects with <6 months of PTNS sessions (i.e., those who exited Part II) will not be enrolled in the 6-month extended follow-up. Those who initiate new FI treatment during Part II are ineligible to enter Part III.

4.3 Study Arm Assignment and Randomization

For Part I, the participant will be randomized to one of the two treatment arms (PTNS or Sham) using a web-based randomization system. The system will supply the site coordinator with a randomization code. Randomization (2:1 to the two treatment arms) will be performed using permuted blocks, with a block size that is known only to the DCC and will be stratified by site and Run-In diary method (phone or paper diary).

For Part II, the respondent participants who elect to continue the study will be randomized to one of two continuing treatment schedules (Fixed or PRN) using a web-based randomization system. The system will supply the site coordinator with a randomization code. Randomization (1:1 to the two treatment arms) will be performed using permuted blocks, with a block size that is known only to the DCC and will be stratified by Part I treatment assignment.

4.4 Masking and Data Lock

4.4.1 General Masking Procedures

Participants will remain masked to Part I treatment assignment throughout the study until at least the completion of the analysis of Part I. After the analysis of Part I and if PTNS shows

improvement over Sham, non-responders from Part I that were assigned to Sham may be invited to re-enroll in the study under the PTNS treatment. These participants will not be masked to PTNS/Sham treatment in Part II.

For Part II, it is not feasible to mask the patients or clinical staff to the treatment schedule due to the nature of how follow-up treatment sessions are scheduled.

4.5 Database Lock

Database lock will occur when data collection has been completed prior to the final data analysis. In order to gain both time and budgetary efficiencies, a database "snapshot" will occur for all analyses performed prior to the final data lock. This will be done to accommodate analysis of the Part I primary outcome prior to the completion of data cleaning for additional outcomes that will be analyzed at a later date. Once the data are saved in the snapshot, subsequent edits to variables in the snapshot can be identified by taking another snapshot and performing a formal comparison of the two data files, reconciling any differences seen and documenting those differences. Until it is agreed upon by the Steering Committee and the DCC that unmasking is not a risk to the study, only the DCC statistician(s) and data manager(s) working directly with the data will be unmasked to the treatment assignments of individual study participants. Thus, data collection of longer-term outcomes will not be compromised by unmasking.

4.6 Part I Study Timeline of Visits and Schedule of Measures

Study Visit Number	Visit 1		Phone Call		Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8	Visit 9	Visit 10	Visit 11	Visit 12	Visit 13	Visit 14	Visit 15
Study Visit Title	Start of Run-In Phase (4 weeks total)	Run-In Diary Week 1	Phone Call Week 3	Run-In Diary Week 4	Baseline	TX 1 (2 weeks after baseline)	TX 2	TX 3	TX 4	TX 5	TX 6	TX 7	TX 8	TX 9	TX 10	TX 11	TX 12	Closure≠
Time in relation to **start of treatment**	-6 weeks	-6 weeks	-4 weeks	-3 weeks	-2 weeks	0	7d	14d	21d	28d	35d	42d	49d	56d	63d	70d	77d	91d
Window	None	None	None	None	2 weeks	1 week	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d
Consent	**X**																	
Assess Eligibility	**X**																	
Pelvic/Rectal Exam	**X**																	
NIDDK Pamphlet	**X**																	
Bowel Diary Instruction	**X**				**X***													
"Run-In" Bowel Diary Recording Begins		**X**		**X**														
Reminder Phone Call to Start Paper Bowel Diary			**X**															
Push Notification to Start Phone Bowel Diary			**X**															
Paper Bowel Diary Collection, Review for Validity & Payment					**X**													
Assess Eligibility to Begin Part I					**X****													
Pregnancy Test					**X*****													
Part I Randomization						**X**												
Blood Specimen Collection						**X********												
Part I Bowel Diary Begins					**X**						**X**						**X**	

Study Visit Number	**Visit 1**		**Phone Call**		**Visit 2**	**Visit 3**	**Visit 4**	**Visit 5**	**Visit 6**	**Visit 7**	**Visit 8**	**Visit 9**	**Visit 10**	**Visit 11**	**Visit 12**	**Visit 13**	**Visit 14**	**Visit 15**
Study Visit Title	Start of Run-In Phase (4 weeks total)	Run-In Diary Week 1	Phone Call Week 3	Run-In Diary Week 4	Baseline	TX 1 (2 weeks after baseline)	TX 2	TX 3	TX 4	TX 5	TX 6	TX 7	TX 8	TX 9	TX 10	TX 11	TX 12	Closure≠
Time in relation to **start of treatment**	-6 weeks	-6 weeks	-4 weeks	-3 weeks	-2 weeks	0	7d	14d	21d	28d	35d	42d	49d	56d	63d	70d	77d	91d
Window	None	None	None	None	2 weeks	1 week	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d
Review Phone Diary Data for Validity & Participant Payment					**X**	**X**							**X**					**X**
PTNS v. Sham Session						**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	
Safety Stimulation Checklist						**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	
AE/SAE						**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Unmasking (Deviation Recorded)						**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Query of Treatment Assignment																		**X**
CRFs/Questionnaires																		
Demographics	**X**																	
Dietary Screener	**X**																	
Physical Exam	**X**																	
PMHx (Update)	**X**				**X**					**X**				**X**				**X**
St. Mark's Score	**X**				**X**					**X**				**X**				**X**
PAC-SYM					**X**													**X**
PISQ-IR					**X**													**X**
Adaptation Index					**X**													**X**
SF-12					**X**													**X**
FIQL					**X**									**X**				**X**
ABLE					**X**									**X**				**X**
PFDI-20					**X**									**X**				**X**

Study Visit Number	**Visit 1**		**Phone Call**		**Visit 2**	**Visit 3**	**Visit 4**	**Visit 5**	**Visit 6**	**Visit 7**	**Visit 8**	**Visit 9**	**Visit 10**	**Visit 11**	**Visit 12**	**Visit 13**	**Visit 14**	**Visit 15**
Study Visit Title	Start of Run-In Phase (4 weeks total)	Run-In Diary Week 1	Phone Call Week 3	Run-In Diary Week 4	Baseline	TX 1 (2 weeks after baseline)	TX 2	TX 3	TX 4	TX 5	TX 6	TX 7	TX 8	TX 9	TX 10	TX 11	TX 12	Closure≠
Time in relation to **start of treatment**	-6 weeks	-6 weeks	-4 weeks	-3 weeks	-2 weeks	0	7d	14d	21d	28d	35d	42d	49d	56d	63d	70d	77d	91d
Window	None	None	None	None	2 weeks	1 week	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d	±3d
PFIQ-7					**X**									**X**				**X**
Modified Manchester/FSFI					**X**									**X**				**X**
PGSC					**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
PGI-I (FI)										**X**				**X**				**X**

≠Participants who exit Part I prematurely will be administered the same panel of questionnaires completed at Visit 15 at the time they exit the study.
*Subjects who did not use the phone app diary in the Run-In phase, will be provided a smartphone and instructed on the Bowel Diary phone app.
** In Part I, a valid diary is defined as 3 consecutive days completed and 10/14 days completed in the 14-day diary.
***As applicable
****Blood collection will be done at Visit 3 for participants who have not been randomized. Blood collection will be done at a time convenient for participants who have already been randomized.

4.7 Part II Study Timeline of Visits and Schedule of Measures

Study Visit Number	**Visit 15**	**Visit 16**	**Visit 17**	**Visit 18**	**Visit 19**	**Visit 20**	**Visit 21**	**Visit 22**	**Visit 23**	**Visit 24**	**Visit 25**	**Visit 26**
Study Visit Title	Part II Randomization Fixed TX 13	Fixed TX 14	Fixed TX 15	Fixed TX 16	Fixed TX 17	Fixed TX 18	Fixed TX 19	Fixed TX 20	Fixed TX 21	Fixed TX 22	Fixed TX 23	Part II Closure*
Time in relation to **first treatment in Part II**	Part II Start Time 0	+2 weeks	+5 weeks	+8 weeks	+12 weeks (6 mo)	+16 weeks	20 weeks	24 weeks (9 mo)	28 weeks	32 weeks	36 weeks	38 weeks (12 mo)
Window	±3d	±3d	±3d	±3d	±7d	±7d	±7d	±7d	±7d	±7d	±7d	±7d
Assess Eligibility for Part II ≠	**X**											
Part II Randomization ≠	**X**											
Bowel Diary Recording Begins					**X**			**X**			**X**	
Review Diary Data for Validity						**X**			**X**			**X**
FIXED Schedule GROUP												
PTNS Treatment	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	
Safety Stimulation Checklist	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	
AE/SAE	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Unmasking (Deviation Recorded)	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	
CRFs/Questionnaires												
PAC-SYM	**X**											**X**
PISQ-IR	**X**											**X**
Adaptation Index	**X**											**X**
SF-12	**X**											**X**
FIQL	**X**				**X**							**X**
ABLE	**X**				**X**							**X**
PFDI-20	**X**				**X**							**X**
PFIQ-7	**X**				**X**							**X**

Study Visit Number	**Visit 15**	**Visit 16**	**Visit 17**	**Visit 18**	**Visit 19**	**Visit 20**	**Visit 21**	**Visit 22**	**Visit 23**	**Visit 24**	**Visit 25**	**Visit 26**
Study Visit Title	Part II Randomization Fixed TX 13	Fixed TX 14	Fixed TX 15	Fixed TX 16	Fixed TX 17	Fixed TX 18	Fixed TX 19	Fixed TX 20	Fixed TX 21	Fixed TX 22	Fixed TX 23	Part II Closure*
Time in relation to **first treatment in Part II**	Part II Start Time 0	+2 weeks	+5 weeks	+8 weeks	+12 weeks (6 mo)	+16 weeks	20 weeks	24 weeks (9 mo)	28 weeks	32 weeks	36 weeks	38 weeks (12 mo)
Window	±3d	±3d	±3d	±3d	±7d	±7d	±7d	±7d	±7d	±7d	±7d	±7d
Modified Manchester/FSFI	**X**				**X**							**X**
PMHx (Update)	**X**				**X**			**X**				**X**
St. Mark's Score	**X**				**X**			**X**				**X**
PGI-I (FI)	**X**				**X**			**X**				**X**
PGSC	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
PRN Schedule GROUP (Visits 15 and 26 must be in-person)												
PTNS Treatment (Projected)	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	
Safety Stimulation Checklist	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	
AE/SAE (at treatment visits or by phone)	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Unmasking (Deviation Recorded)	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	
CRFs/Questionnaires**												
PAC-SYM	**X**											**X**
PISQ-IR	**X**											**X**
Adaptation Index	**X**											**X**
SF-12	**X**											**X**
FIQL	**X**				**X**							**X**
ABLE	**X**				**X**							**X**
PFDI-20	**X**				**X**							**X**
PFIQ-7	**X**				**X**							**X**
Modified Manchester/FSFI	**X**				**X**							**X**
PMHx update (visit or by phone)	**X**				**X**			**X**				**X**

Study Visit Number	**Visit 15**	**Visit 16**	**Visit 17**	**Visit 18**	**Visit 19**	**Visit 20**	**Visit 21**	**Visit 22**	**Visit 23**	**Visit 24**	**Visit 25**	**Visit 26**
Study Visit Title	Part II Randomization Fixed TX 13	Fixed TX 14	Fixed TX 15	Fixed TX 16	Fixed TX 17	Fixed TX 18	Fixed TX 19	Fixed TX 20	Fixed TX 21	Fixed TX 22	Fixed TX 23	Part II Closure*
Time in relation to **first treatment in Part II**	Part II Start Time 0	+2 weeks	+5 weeks	+8 weeks	+12 weeks (6 mo)	+16 weeks	20 weeks	24 weeks (9 mo)	28 weeks	32 weeks	36 weeks	38 weeks (12 mo)
Window	±3d	±3d	±3d	±3d	±7d	±7d	±7d	±7d	±7d	±7d	±7d	±7d
St. Mark's Score	**X**				**X**			**X**				**X**
PGI-I (FI)	**X**				**X**			**X**				**X**
PGSC*** (at visits or by phone)	**In-person***	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**In-person***

≠An improvement in St. Mark's ≥ 4 from the participant's baseline
*All participants will attend Visit 26. Those who exit Part II prematurely will be administered the Visit 26 panel of questionnaires at the time they exit the study.
**The PRN group will complete the questionnaires electronically via an email link at the 6 and 9-month time point if they do not have a treatment visit at 6 months and/or 9 months. They will attend Visit 26 in-person and will complete the questionnaires in-person. Mailing or administering paper questionnaires should be used only as a back-up. Other planned assessments (e.g., AE collection and medical history update) will be done by phone if the participant does not attend a treatment visit in-person.
***The PGSC will be administered to participants in the PRN group in-person at Visits 15 and 26. At all other timepoints, the coordinator will administer the PGSC by phone if the participant does not schedule an in-person treatment.

4.8 Part III Study Timeline of Visits and Schedule of Measures

Schedule of measure to be completed online through secure email link						
Weeks after last PTNS treatment	**6 weeks**	**10 weeks**	**14 weeks**	**18 weeks**	**22 weeks**	**26 weeks Part III EXIT**
PGSC by phone	**X**	**X**	**X**	**X**	**X**	**X**
ePRO						
St. Mark's Score	**X**	**X**	**X**	**X**	**X**	**X**
PGI-I (FI)	**X**	**X**	**X**	**X**	**X**	**X**
FIQL	**X**	**X**	**X**	**X**	**X**	**X**
Adaptation Index	**Only if PGSC ≤2**					**X**

5 ANALYSIS POPULATIONS

5.1 Intention-to-Treat (ITT) Population

The primary analysis population and the population for all secondary analyses will be the intention to treat population, which includes all randomized and eligible subjects. All subjects will be assigned to the arm to which they were randomized irrespective of treatment received. The analysis will include randomized patients who provided any post-treatment outcome data.

5.2 Per-Protocol (PP) Population

Per-protocol analyses (in which participants are analyzed according to the treatment actually received and their degree of compliance with those treatments) will be considered exploratory secondary analyses. Participants who were adherent to their assigned treatment will be included in the per-protocol analysis. For Part I, adherence to the study regimen will be defined as completing 10 of the 12 stimulation sessions within a 14-week window. For Part II, adherence will be defined as attendance at 9 of 11 sessions for those assigned to the fixed schedule

5.3 Safety (SAF) Population

The safety population will comprise all subjects who received any study treatment grouped by randomized treatment, irrespective of treatment received or eligibility.

6 SAMPLE SIZE DETERMINATION

This randomized, single-masked controlled trial is designed to evaluate the effect of PTNS for FI by comparing the effectiveness of PTNS to a validated sham in women with FI. The St. Mark's (Vaizey) score has been selected as the primary outcome measure. This decision was based upon the information collected as a part of the CAPABLe study design activity. This sample size estimate has taken into consideration the background literature on properties of the Vaizey scale as utilized in the CAPABLe design and the published literature on use of PTNS for FI. Key information that is relevant to the NOTABLe study design is summarized in bullet form below:

- While several small studies on PTNS have been conducted, the measure most frequently used for those studies was the Wexner score rather than the St. Mark's (Vaizey) score.
- The MID for the St. Mark's (Vaizey) scale for FI selected for the CAPABLe study was a difference between arms of 5 units, based on Bols (58). While the information on the effect of PTNS based on the Vaizey scale is limited, the effect sizes found to date are smaller than this difference. Given the limited data available, estimates were developed that would allow us to detect effect sizes ranging from half of the MID to the MID.
- While very limited information is available, the work done by Vaizey during the original development of the scale indicates that the underlying variability of the Wexner scale is comparable to the variability of the St. Mark's (Vaizey) scale (standard deviation of 5.9 versus a standard deviation of 6). Consequently, while we can't use information on the Wexner directly for this study, the information presented in Table 1 that shows the relative consistency of variability before and after treatment with the Wexner and the consistency of the variability of the Wexner and the Vaizey scores provides some indication that the preliminary estimates of variability should be relatively robust.
- Three manuscripts (Bols, 2010, de la Portilla, 2009, and Queralto, 2006) provide sufficient information to estimate the correlation of measures before and after treatment, although the treatment for the Queralto study is transcutaneous PTNS. Preliminary calculations indicate that the correlation can be estimated conservatively to be about 0.35.

- Although the original Vaizey manuscript indicated that the standard deviation of the outcome measure is around 6, the other studies available suggest that that standard deviation is at the high end of the range. At the time of NOTABLE sample size calculation, unpublished data from CAPABLe indicated that the standard deviation for the change from baseline in St. Mark's score at 12 weeks among women assigned to education and placebo drug was close to 7. Based on the range of available data, we have elected to conservatively design the trial for a standard deviation of 7.

Based on the results outlined above, preliminary sample size estimates were generated under the assumption that the effect size of interest (difference in the change from baseline in Vaizey score at the end of the treatment period between the PTNS arm and the sham arm) was in the range of 2.5 to 5, and that the standard deviation was 4.5, 5, 6, or 7. The calculations assumed that we wanted the study to achieve a power of 90% and the outcome measure of interest is the change from baseline in Vaizey score at the end of the treatment period. The results of the preliminary calculations are shown in Table 2.

Because one interest of the study is to evaluate alternative maintenance strategies for the PTNS treatment, calculations were generated under the assumption that the ratio of randomization of PTNS to sham participants was 2:1. The sample sizes in the table represent the total sample size for both the treatment and sham arms.

The protocol committee elected a mid-range effect size of 4.0 accepting that it was less than the MID of 5.0. Based on these assumptions, a minimum of 147 women (98 randomized to PTNS and 49 to sham) would need to have outcomes assessed after 12 weeks of stimulation to have 90% power to detect a significant difference between groups in change from baseline in St. Mark's (Vaizey) score using a two-sided test with an alpha level of 0.05. To account for a potential 10% drop out rate during Part I, the sample size has been inflated to 165 (110 assigned to PTNS and 55 to sham).

Effect Size (Difference in PTNS and Sham Score at Study End)	Total Sample Size Needed to Achieve 90% Power			
	SD=4.5	SD=5	SD=6	SD=7
2.5	156	192	276	372
3.0	111	135	192	261
3.5	81	99	141	192
4.0	63	78	111	147
4.5	51	63	87	117
5.0	42	51	72	96

In the Overactive Bladder Innovative Therapy (OrBIT) study of PTNS vs. Detrol for UI, 70% of participants randomized to PTNS were responders at 3 months. We conservatively assume that 60% of NOTABLe participants in the PTNS group will be classified as responders at the end of Part I and that the remaining 40% will be non-responders and/or will drop out of the study.

The goal of Part II is to estimate the percent of women assigned to each maintenance schedule who are still responders at one year from the start of PTNS treatment. In OrBIT, approximately 70% of participants who began Part II were responders at one year. Making a conservative assumption that the Part II responder rate in NOTABLe will be 50%, the table below shows the number of Part II

participants needed in each maintenance group for the 95% confidence interval (CI) around a 50% responder rate in that group to have a half-width of 10% to 15%.

95% CI half-width	10%	11%	12%	13%	14%	15%
Part II participants	97	80	67	57	49	43

The protocol team decided that a 95% confidence interval half-width of 15% would provide adequate information about the PTNS maintenance strategies for planning a future study, thus requiring 86 PTNS responders to be enrolled in Part II (43 assigned to each maintenance strategy). To achieve that number under the previous assumption of a 60% responder rate after 12 weeks of stimulation, and assuming 90% of Part I responders enroll in Part II, an estimated 178 women need to begin PTNS treatment in order for 86 to be randomized to Part II.

In order to randomize 110 participants to PTNS and 55 to sham in Part I of the study, we anticipate needing to enroll 254 women in the Run-In Phase, assuming that 165 (65%) will be eligible to go on to Part I after the Run-In Phase. We expect that 53 women randomized to PTNS in Part I will go on to Part II, assuming that 59/98 (60%) of those who complete Part I will respond to PTNS and 53 (90% of responders) will enter Part II. Thus, an additional 33 PTNS responders will need to be randomized to a maintenance strategy in order to meet the Part II randomization target of 86 participants. We expect that an additional 94 women will need to be enrolled in the Run-In Phase in order to complete Part II. This assumes that 68 (65%) will be eligible to begin PTNS after the Run-In Phase, 7 (10%) will be lost to follow up, 37 (60%) will be PTNS responders, and 33 (90% of responders) will be randomized to a maintenance strategy in Part II. Thus, a total of 359 participants will be enrolled in the Run-In Phase in order to complete both Part I and Part II.

All subjects (PTNS and sham) randomized to Part II (maintenance) and with ≥6 months total of assigned stimulation sessions will be asked to provide symptom updates through the extended 6-month follow-up of symptom control, improvement, and QOL after final PTNS treatment. Those who initiate new FI treatment during Part II are ineligible to enter Part III. Sham subjects will also be asked for symptom updates to maintain masking of all subjects. Subjects who completed Part II prior to approval of the protocol amendment for extended follow-up (Part III) will be contacted by study staff and invited to re-enter follow-up under an IRB approved consent amendment. A total of 86 subjects assigned to PTNS will be randomized to a Part II and based upon a 30% responder rate in the sham group up to 17 sham subjects will provide this extended post-stimulation follow-up data. Data will be analyzed from only those subjects who received PTNS.

7 STATISTICAL AND ANALYTICAL ISSUES

7.1 General Rules

All statistical computations will be performed and data summaries will be created using SAS 9.4 or higher. If additional statistical packages are required, these will be discussed in the study report. For summaries of study data, categorical measures will be summarized in tables listing the frequency and the percentage of subjects in each study arm; continuous data will be summarized by presenting mean, standard deviation, median and range; and ordinal data will be summarized by only presenting median and range.

7.2 Adjustments for Covariates

Indicator variables for the study stratification of site will be included as covariates in most efficacy analyses performed for this study (details in section 9). Additionally, demographic and baseline characteristics for subjects will be compared between study arms using analysis of covariance techniques for continuous measures, Mantel-Haenszel mean score test using standardized midrank scores for ordinal measures, and Cochran Mantel-Haenszel chi-square tests for general association for categorical measures. If sample sizes allow, these analyses will control for the study stratification factors. If these analyses suggest that substantial differences exist among arms, the use as covariates of these parameters on which the arms differ will be explored in secondary exploratory analyses of the efficacy data.

7.3 Handling of Dropouts and Missing Data

Missing data mechanisms will be explored, and sensitivity analyses will be conducted on primary outcomes to assess the robustness of the described analyses. Methods employed for sensitivity analyses may include multiple imputation or inverse probability weighting methodology. Imputation is not planned for secondary analyses.

7.4 Interim Analyses and Data Monitoring

Safety outcomes will be assessed at each DSMB meeting. Rates of safety outcomes will be compared between treatment groups using Fisher's exact tests and provided to the DSMB. There will be no formal interim analyses of efficacy outcomes. At each meeting, the DSMB will be presented with information about enrollment and outcome data attainment (for example, the percent of expected visits that have been completed) to allow them to determine that the study is making reasonable progress.

At the completion of Part I, an analysis of the primary outcome (St. Mark's (Vaizey) Score) will be performed in order to determine if PTNS is significantly more effective than Sham at reducing fecal incontinence symptoms. The analysis will be conducted using the statistical model described in Section 9.3, and a p-value less than or equal to 0.05 will be considered statistically significant.

In the event that PTNS is found to be more effective than Sham as described above, the study will continue as planned in the protocol. Part II Sham participants will be notified that their treatments will be discontinued and Part I enrollment will be re-opened without randomized treatment in order to obtain the target sample size for Part II. If PTNS is not found to be more effective than Sham, the study will be halted, all Part II participants will be notified that all treatments will be discontinued, and the extended follow-up (Part III) will be stopped.

7.5 Masked Data Review

A masked data review of the primary outcome and secondary outcomes for this study will be performed by the protocol team. This review will occur prior to completion of the Part I analyses. This will include a presentation of descriptive statistics (e.g. means, standard deviations, percentiles for continuous variables and counts and percentages of categorical variables) of the selected outcomes and model predictor variables.

7.6 Multicenter Studies

For this multicenter study, randomization of study participants was stratified within center. Consequently, for all model-based primary and secondary analyses, center will be included as a fixed effect in the models. As an ancillary analysis associated with the primary outcome we will examine descriptively whether the treatment effect varies across sites; however, no other analyses

will assess site differences in treatment effect because sample sizes are inadequate to support evaluation of site-level effects.

7.7 Multiple Comparisons and Multiplicity

The Part I primary hypothesis will be tested at a nominal two-sided type I error of 0.05. All p-values for any baseline and demographic characteristic comparisons, secondary outcomes, and safety parameters will be for descriptive purposes only.

7.8 Examination of Subgroups

No subgroup analyses are planned.

7.9 Assessment Windows

For the primary analysis, decisions about how to treat out-of-window visits will be made prior to unmasking data. For secondary analyses, all available data will be used.

8 STUDY SUBJECT CHARACTERIZATION

8.1 Subject Disposition

Participant eligibility status will be summarized and listed by study arm and overall disposition of study participants will be described using a standard cohort diagram. The number of subjects randomized; completing or discontinuing from study therapy; completing each follow-up visit will be summarized by study arm. Reasons for study treatment discontinuation and study withdrawal will be listed.

8.2 Protocol Deviations

Protocol deviations are identified via automated checks of the clinical database and reported by site study coordinators in the study data management system. Protocol deviations will be listed by site with information such as type of deviation, time of occurrence, and reason. Incidence rate of protocol deviations will also be summarized overall and for each protocol deviation category by site. Incidence rate of protocol deviations will be calculated as: number of deviations divided by the number of subject months at the site

8.3 Demographic and Baseline Characteristics

Demographic and baseline clinical characteristics for the study participants will be summarized by study arm using the general analysis rules describe above. Variables of interest include: age (years), parity, gravidity, race and ethnicity, marital status, education level (classified as binary variable as having some college or greater or no college education), health insurance status (private only, Medicare/Medicaid only, combination of both), smoking status (never, previous, current), prior anal/rectal surgery, BMI, and baseline levels of all QOL measures.

9 EFFICACY ANALYSES

9.1 Overview of Efficacy Analyses Methods

- All efficacy analyses will be performed on the ITT population.
- For Part I, all efficacy variables will be summarized by treatment group at baseline and at Part I time points at which assessments were administered and collected. For Part II, all efficacy variables will be summarized by treatment group at Visit 15 (Part II baseline) and at the Visit 19, Visit 22, and Visit 26 time points. N, mean, standard deviation,

minimum, and maximum will summarize continuous efficacy variables, whereas number and percent will summarize categorical efficacy variables.

- Unless otherwise noted, all analyses of dichotomous outcomes, measured at respective endpoints, will be performed using a generalized linear mixed model. Part I models will be adjusted for stratification by clinical site. If there are not enough patients in every clinical site to include the variable in the models as a fixed effect, clinical site will be included as random intercepts to account for correlation between outcomes of patients treated by the same clinical site or similar sites will be combined to reduce small group sizes. Consistent with the description of the primary analysis in the protocol, analyses of primary and secondary outcomes will also include an independent variable for time. For analysis of measures assessed at multiple time points (for example, Visit 7 and Visit 11), longitudinal modeling will be used and the interaction between time and treatment groups will be included. Unless otherwise noted, all analyses of continuous efficacy variables (e.g., QOL scales) will be performed using general linear mixed models. Variables with distributions substantially different from normal will be transformed prior to analysis. Models will be adjusted for clinical site. If there are not enough patients per clinical site to include the variables in the models as fixed effects, clinical site will be included as a random intercept to account for correlation between outcomes of patients treated by the same clinical site or similar sites will be combined. For analysis of measures assessed at multiple time points, longitudinal modeling will be used and the interaction between time and treatment groups will be included. Under the assumption that any missing outcome data will be missing at random (thus, missing St. Mark's (Vaizey) scores at Visit 15 may be related to earlier outcomes and covariates), this model will produce more accurate estimates in the presence of missing data than one that models only outcomes at Visit 15. For the primary outcome and for secondary outcomes, models will include 2-way interactions between treatment assignments and time.

9.2 Efficacy Variables

Primary and secondary efficacy variables as well as exploratory and safety outcomes are described in the table below.

Variable	Type	Definition
Primary Outcomes		
Part I: Change from baseline in St. Mark's (Vaizey) Score through Visit 15	Continuous	The Vaizey bowel incontinence score will be computed at baseline and at each visit during which assessments are administered and collected using the standard scoring algorithm. Specifically, sum responses across all questions (Q1-Q4: 0=never, 1=rarely, 2=sometimes, 3=weekly, 4=daily; Q5-Q6: 0=no, 2=yes; Q7: 0=no, 4=yes). If any responses are missing then no score will be calculated. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing. This may be calculated inside the DMS.
Part II: Responder Status at 12 Months	Dichotomous (Yes/No)	Responder Status=Yes is defined as ≥ 4-point reduction from baseline St. Mark's (Vaizey) score. This is calculated only for Part II participants who were assigned to PTNS; Sham group participants will be excluded from Part II analysis. Responder status will be calculated based on data from the 12-month visit, and participants who drop out of Part II prior to the 12-month visit will be considered non-responders for analysis purposes.
Secondary Outcomes		
Change from baseline in Fecal Incontinence Episodes (FIEs) through Visit 15	Continuous	The total number of FIEs will be taken from the bowel diary. *See Attachment 1.* The outcome will then be computed as the difference in number of FIEs at each post-baseline time point the diary is collected and the number of FIEs at baseline. If data for the assessment time point are missing, the outcome variable will be coded as missing.

Variable	Type	Definition
Change from baseline in Urgency Episodes through Visit 15	Continuous	The total number of urgency episodes will be taken from the bowel diary. *See Attachment 1.* The outcome will then be computed as the difference in number of urgency episodes at each post-baseline time point the diary is collected and the number of urgency episodes at baseline. If data for the assessment time point are missing, the outcome variable will be coded as missing.
Change from baseline in Bowel Movements (BMs) through Visit 15	Continuous	The total number of BMs will be taken from the bowel diary. *See Attachment 1.* The outcome will then be computed as the difference in number of BMs at each post-baseline time point the diary is collected and the number of BMs at baseline. If data for the assessment time point are missing, the outcome variable will be coded as missing.
Change from baseline in ABLe Liquid Stool Score through Visit 15	Continuous	The ABLe Liquid Stool Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, calculate the average of the non-missing responses for questions A1 (4=large amount; 3=moderate amount, 2=small amount; 1=staining only, 0=none), A2 (4=daily; 3=weekly; 2=monthly; 1=less than monthly; 0=never), and A3 (4=very bothered; 3=somewhat bothered; 2=a little bothered; 1=not at all bothered; 0=not applicable). The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in ABLe Solid Stool Score through Visit 15	Continuous	The ABLe Solid Stool Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, calculate the average of the non-missing responses for questions A4 (4=large amount; 3=moderate amount, 2=small amount; 1=staining only, 0=none), A5 (4=daily; 3=weekly; 2=monthly; 1=less than monthly; 0=never), and A6 (4=very bothered; 3=somewhat bothered; 2=a little bothered; 1=not at all bothered; 0=not

Variable	Type	Definition
		applicable). The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in ABLe Mucus Score through Visit 15	Continuous	The ABLe Mucus Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, calculate the average of the non-missing responses for questions A7 (4=large amount; 3=moderate amount, 2=small amount; 1=staining only, 0=none), A8 (4=daily; 3=weekly; 2=monthly; 1=less than monthly; 0=never), and A9 (4=very bothered; 3=somewhat bothered; 2=a little bothered; 1=not at all bothered; 0=not applicable). The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in ABLe Gas Score through Visit 15	Continuous	The ABLe Gas Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, calculate the average of the non-missing responses for questions A10 (4=daily; 3=weekly; 2=monthly; 1=less than monthly; 0=never) and A11 (4=very bothered; 3=somewhat bothered; 2=a little bothered; 1=not at all bothered; 0=not applicable). The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in ABLe Predictability/Awareness Score through Visit 15	Continuous	The ABLe Predictability/Awareness Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, calculate the average of the non-missing responses for questions A12 (0.8=all of the time; 1.6=most of the time; 2.4=some of the time; 3.2=a little of the time; 4=none of the time; 0=not applicable) and A13 (4=almost always; 3.2=often; 2.4=sometimes; 1.6=rarely; 0.8=never; 0=Not applicable). The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.

Variable	Type	Definition
Change from baseline in ABLe Control Score through Visit 15	Continuous	The ABLe Control Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, calculate the average of the non-missing responses for questions A14 (4=almost always; 3=often; 2=sometimes; 1=rarely; 0=never) and A15 (4=very bothered; 3=somewhat bothered; 2=a little bothered; 1=not at all bothered; 0=not applicable). The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in ABLe Ancillary Bowel Symptoms Score through Visit 15	Continuous	The ABLe Ancillary Bowel Symptoms Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, calculate the average of the non-missing responses for questions A16 (0=almost always; 1=often; 2=sometimes; 3=rarely; 4=never), A17 (4=very bothered; 3=somewhat bothered; 2=a little bothered; 1=not at all bothered; 0=not applicable), ad B1 (4=almost always; 3=often; 2=sometimes; 1=rarely; 0=never). The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in ABLe Overall Score through Visit 15	Continuous	The ABLe Overall Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, calculate the average of the non-missing responses for questions A1 through A17 and question B1 (individual item scores as noted previously). The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in FIQL Lifestyle Score through Visit 15	Continuous	The FIQL Lifestyle Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, calculate the average of the non-missing responses for questions 2a, 2b, 2c, 2d, 2e, 2g, 2h (1=most of the time; 2=some of the time; 3=little of the time; 4=none of the time) and questions 3b, 3l, and 3m (1=strongly agree; 2=somewhat agree; 3=somewhat disagree; 4=strongly disagree). The outcome will then be computed as the

Variable	Type	Definition
		difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in FIQL Coping/Behavior Score through Visit 15	Continuous	The FIQL Coping/Behavior Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, calculate the average of the non-missing responses for questions 2f, 2i, 2j, 2k, 2m (1=most of the time; 2=some of the time; 3=little of the time; 4=none of the time) and questions 3c, 3h, 3j, and 3n (1=strongly agree; 2=somewhat agree; 3=somewhat disagree; 4=strongly disagree). The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in FIQL Depression/Self Perception Score through Visit 15	Continuous	The FIQL Depression/Self Perception Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, calculate the average of the non-missing responses for question Q1 (1=poor; 2=fair; 3=good; 4=very good; 5=excellent), questions 3d, 3f, 3g, 3i, and 3k (1=strongly agree; 2=somewhat agree; 3=somewhat disagree; 4=strongly disagree), and question 4 (1=extremely so; 2=very much so; 3=quite a bit; 4=some – enough to bother me; 5=a little bit; 6=not at all). The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in FIQL Embarrassment Score through Visit 15	Continuous	The FIQL Embarrassment Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, calculate the average of the non-missing responses for question 2l (1=most of the time; 2=some of the time; 3=little of the time; 4=none of the time) and questions 3a and 3e (1=strongly agree; 2=somewhat agree; 3=somewhat disagree; 4=strongly disagree). The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.

Variable	Type	Definition
Change from baseline in MMHQ Incompetence Impact Score through Visit 15	Continuous	The MMHQ Incontinence Impact subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, subtract 1 from the response to Q1, divide by 4 and multiply by 100. If the response to Q1 is missing, no score is calculated. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing..
Change from baseline in MMHQ Role Limitations Score through Visit 15	Continuous	The MMHQ Role Limitations subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, subtract 1 from each of the responses to Q6 and Q7, then sum the responses, divide by 8, and multiply by 100. If the response to either Q6 or Q7 is missing, no score is calculated. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in MMHQ Physical Limitations Score through Visit 15	Continuous	The MMHQ Physical Limitations subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, subtract 1 from each of the responses to Q8 and Q9, then sum the responses, divide by 8, and multiply by 100. If the response to either Q8 or Q9 is missing, no score is calculated. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in MMHQ Social Limitations Score through Visit 15	Continuous	The MMHQ Social Limitations subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, subtract 1 from each of the responses to Q10, Q11, and Q13, then sum the responses, divide by 12, and multiply by 100. If the response to Q13 is missing, then divide by 8 instead of 12. If the response to either Q10 or Q11 is missing, no score is calculated. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.

Variable	Type	Definition
Change from baseline in MMHQ Personal Relationships Score through Visit 15	Continuous	The MMHQ Personal Relationships subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, subtract 1 from each of the responses to Q12 and Q20, then sum the responses, divide by 8, and multiply by 100. If the response to either Q12 or Q20 is missing, then divide by 4 instead of 8. If the response to both Q12 and Q20 is missing, no score is calculated. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in MMHQ Emotions Score through Visit 15	Continuous	The MMHQ Emotions subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, subtract 1 from each of the responses to Q14, Q15, and Q16, then sum the responses, divide by 12, and multiply by 100. If the response to Q14, Q15, or Q16 is missing, no score is calculated. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in MMHQ Sleep/Energy Score through Visit 15	Continuous	The MMHQ Sleep/Energy subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, subtract 1 from each of the responses to Q17 and Q18, then sum the responses, divide by 8, and multiply by 100. If the response to Q17 or Q18 is missing, no score is calculated. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in MMHQ Severity Meaures Score through Visit 15	Continuous	The MMHQ Severity Measures subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, subtract 1 from each of the responses to Q24, Q25, Q26, Q27, and Q28, then sum the responses, divide by 20, and multiply by 100. If the response to Q24, Q25, Q26, Q27, or Q28 is missing, no score is calculated. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the

<table>
<tr><th>Variable</th><th>Type</th><th>Definition</th></tr>
<tr><td></td><td></td><td>score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.</td></tr>
<tr><td>Change from baseline in FISI Patient Score through Visit 15</td><td>Continuous</td><td>The FISI Patient Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the weighted scores for Q2, Q3, Q4, and Q5 where weights are assigned as follows:
<table>
<tr><th rowspan="2">Question</th><th colspan="6">Response Weights</th></tr>
<tr><th>Never</th><th>1-3 times a month</th><th>Once a week</th><th>2 or more times a week</th><th>Once a day</th><th>2 or more times a day</th></tr>
<tr><td>Q2. Solid</td><td>0</td><td>8</td><td>10</td><td>13</td><td>16</td><td>18</td></tr>
<tr><td>Q3. Liquid</td><td>0</td><td>8</td><td>10</td><td>13</td><td>17</td><td>19</td></tr>
<tr><td>Q4. Mucus</td><td>0</td><td>3</td><td>5</td><td>7</td><td>10</td><td>12</td></tr>
<tr><td>Q5. Gas</td><td>0</td><td>4</td><td>6</td><td>8</td><td>11</td><td>12</td></tr>
</table>
If there are missing responses, then no score is calculated. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.</td></tr>
<tr><td>Change from baseline in FISI Doctor Score through Visit 15</td><td>Continuous</td><td>The FISI Doctor Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the weighted scores for Q2, Q3, Q4, and Q5 where weights are assigned as follows:
<table>
<tr><th rowspan="2">Question</th><th colspan="6">Response Weights</th></tr>
<tr><th>Never</th><th>1-3 times a month</th><th>Once a week</th><th>2 or more times a week</th><th>Once a day</th><th>2 or more times a day</th></tr>
<tr><td>Q2. Solid</td><td>0</td><td>11</td><td>14</td><td>16</td><td>17</td><td>19</td></tr>
<tr><td>Q3. Liquid</td><td>0</td><td>10</td><td>13</td><td>14</td><td>16</td><td>18</td></tr>
<tr><td>Q4. Mucus</td><td>0</td><td>5</td><td>7</td><td>7</td><td>9</td><td>11</td></tr>
</table></td></tr>
</table>

Variable	Type	Definition
		<table><tr><td>Q5. Gas</td><td>0</td><td>2</td><td>4</td><td>6</td><td>8</td><td>9</td></tr></table> If there are missing responses, then no score is calculated. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in PFDI-20 UDI subscale through Visit 15	Continuous	The PFDI-20 UDI subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, average the responses for questions Q15 through Q20 (0=no, 1=yes: not at all, 2=yes: somewhat, 3=yes: moderately, 4=yes: quite a bit) and multiply by 25. If there are missing responses, then the score is the average of the non-missing responses multiplied by 25. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in PFDI-20 CRADI subscale through Visit 15	Continuous	The PFDI-20 CRADI subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, average the responses for questions Q7 through Q14 (0=no, 1=yes: not at all, 2=yes: somewhat, 3=yes: moderately, 4=yes: quite a bit) and multiply by 25. If there are missing responses, then the score is the average of the non-missing responses multiplied by 25. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in PFDI-20 POPDI subscale through Visit 15	Continuous	The PFDI-20 POPDI subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, average the responses for questions Q1 through Q6 (0=no, 1=yes: not at all, 2=yes: somewhat, 3=yes: moderately, 4=yes: quite a bit) and multiply by 25. If there are missing responses, then the score is the average of the non-missing responses multiplied by 25. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.

Variable	Type	Definition
Change from baseline in PFDI-20 Global Score through Visit 15	Continuous	The PFDI-20 Global Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the scores from the UDI, CRADI, and POPDI subscales. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in PFIQ-7 UIQ subscale through Visit 15	Continuous	The PFIQ-7 UIQ subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, average the responses for questions Q1a through Q7a (0=not at all, ... 3=quite a bit) and multiply by 100/3. If there are missing responses, then the score is the average of the non-missing responses multiplied by 100/3. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in PFIQ-7 CRAIQ subscale through Visit 15	Continuous	The PFIQ-7 CRAIQ subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, average the responses for questions Q1b through Q7b (0=not at all, ..., 3=quite a bit) and multiply by 100/3. If there are missing responses, then the score is the average of the non-missing responses multiplied by 100/3. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in PFIQ-7 POPIQ subscale through Visit 15	Continuous	The PFIQ-7 POPIQ subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, average the responses for questions Q1c through Q7c (0=not at all, …, 3=quite a bit) and multiply by 100/3. If there are missing responses, then the score is the average of the non-missing responses multiplied by 100/3. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.

Variable	Type	Definition
Change from baseline in PFIQ-7 Global Score through Visit 15	Continuous	The PFIQ-7 Global Score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the scores from the UIQ, CRAIQ, and POPIQ subscales. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in Defecatory symptoms as measured by the Patient Assessment of Constipation Symptoms (PAC-SYM) questionnaire through Visit 15	Continuous	The PAC-SYM score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms Specifically, sum responses across all questions (0=absent, 1=mild, 2=moderate, 3=severe, 4=very severe). If there are any missing responses, then a total score is not calculated. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at a time point for the assessment are missing, the outcome variable will be coded as missing.
Change from baseline in PISQ-IR not sexually active – partner related subscale score through Visit 15	Continuous	The PISQ-IR not sexually active – partner related subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the scores for questions Q2a, Q2b (1=strongly agree, …, 4=strongly disagree). If there is more than 1 missing response, then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at any time point are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are not sexually active.
Change from baseline in PISQ-IR not sexually active – condition specific subscale score through Visit 15	Continuous	The PISQ-IR not sexually active – condition specific subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the scores for questions Q2c, Q2d, Q2e (1=strongly agree, …, 4=strongly disagree). If there is more than 1 missing response, then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at 12 and 24 weeks and the score at baseline. If data

Variable	Type	Definition
		at any time point are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are not sexually active.
Change from baseline in PISQ-IR not sexually active – global quality subscale score through Visit 15	Continuous	The PISQ-IR not sexually active – global quality subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the scores for questions Q4a, Q4b, Q5a, and Q6 using reverse scores for all but Q5a (Q4a and Q4b are Likert scales of 1 to 5; Q5a: 1=strongly agree, …, 4=strongly disagree; Q6: 1=not at all, …, 4=a lot). If there are more than 2 missing responses, then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at any time point are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are not sexually active.
Change from baseline in PISQ-IR not sexually active – condition impact subscale score through Visit 15	Continuous	The PISQ-IR not sexually active – condition subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the scores for questions Q3, Q5b, Q5c using reverse scores for Q3 (Q3: 1=not at all, …, 4=a lot; Q5b, Q5c: 1=strongly agree, …, 4=strongly disagree). If there is more than 1 missing response, then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at any time point are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are not sexually active.
Change from baseline in PISQ-IR sexually active – arousal, orgasm subscale score through Visit 15	Continuous	The PISQ-IR sexually active – arousal, orgasm subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the scores for questions Q7, Q8a, Q10, and Q11 using reverse scores for Q11 (Q7, Q8a, Q11: 1=never, ..., 5=[almost] always; Q10: 1=much less intense, ..., 5=much more intense; check box response to Q1=1). If there are more than 2 missing responses, then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items

Variable	Type	Definition
		answered. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at any time point are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are sexually active.
Change from baseline in PISQ-IR sexually active – condition specific subscale score through Visit 15	Continuous	The PISQ-IR sexually active – condition specific subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the scores for questions Q8b, Q8c, Q9 using reverse scores for all (Q8b, Q8c, Q9: 1=never, ..., 5=[almost] always). If there is more than 1 missing response, then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at any time point are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are sexually active.
Change from baseline in PISQ-IR sexually active – partner related subscale score through Visit 15	Continuous	The PISQ-IR sexually active – partner related subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the scores for questions Q13, Q14a, Q14b using reverse scores for Q14a and Q14b (Q13: 1=all of the time, ..., 4=hardly ever/rarely; Q14a, Q14b: 1=very positive, ..., 4=very negative). If there is more than 1 missing response, then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at any time point are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are sexually active and have a sexual partner.
Change from baseline in PISQ-IR sexually active – desire subscale score through Visit 15	Continuous	The PISQ-IR sexually active – desire subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the scores for questions Q15, Q16, Q17 using reverse scores for Q16 and Q17 (Q15: 1=never, ..., 5=always; Q16: 1=daily, ...,5=never; Q17: 1=very high, ..., 5=very low or none at all). If there is more than 1 missing

Variable	Type	Definition
		response, then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at any time point are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are sexually active.
Change from baseline in PISQ-IR sexually active – condition impact subscale score through Visit 15	Continuous	The PISQ-IR sexually active – condition impact subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the scores for questions Q18, Q20b-d using reverse scores for Q18 (Q18: 1=not at all, ..., 4=a lot; Q20b-d: 1=strongly agree, ..., 4=strongly disagree). If there are more than 2 missing responses, then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at any time point are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are sexually active.
Change from baseline in PISQ-IR sexually active – global quality rating subscale score through Visit 15	Continuous	The PISQ-IR sexually active – global quality rating subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the scores for questions Q19a-Q19c, Q20a using reverse scores for Q19a-Q19c (Q19a-c: 1=satisfied, ..., 5=dissatisfied; Q20a: 1=strongly agree, ..., 4=strongly disagree). If there are more than 2 missing responses, then a total score is not calculated. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at any time point are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are sexually active.

Variable	Type	Definition
Change from baseline in PISQ-IR sexually active – average score through Visit 15	Continuous	The PISQ-IR sexually active – average score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, sum the scores for questions Q7, Q8a, Q8b, Q8c, Q9, Q10, Q11, Q13, Q14a, Q14b, Q15, Q16, Q17, Q18, Q19a, Q19b, Q19c, Q20a, Q20b, Q20c, and Q20d using reverse scores for Q8b, Q8c, Q9, Q11, Q14a, Q14b, Q16, Q17, Q18, Q19a, Q19b, and Q19c. Point values have been previously noted. If there are more than 10 missing responses, then a total score is not calculated. If the subject does not have a sexual partner (Q12=No), then Q13, Q14a, and Q14b are removed from the sum and there must be at least 9 non-missing responses to calculate a score. To handle missing values, the final score is obtained by dividing the sum by the number of items answered. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at any time point are missing, the outcome variable will be coded as missing. Scores should only be calculated for participants that are sexually active.
Change from Baseline Fecal Incontinence Adaptation Index Hygiene subscale through Visit 15	Continuous	The AI Hygiene subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, average the responses for questions Q1-Q4, Q9, and Q14 (0=never, 25=rarely, 50=sometimes, 75=often, 100=always). If there are at least 5 non-missing responses, then the score is the average of the non-missing responses, otherwise, the score is not calculated. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at any time point are missing, the outcome variable will be coded as missing.
Change from Baseline Fecal Incontinence Adaptation Index Avoidance subscale through Visit 15	Continuous	The AI Avoidance subscale score will be computed at baseline and at each visit during which assessments are administered and collected using standard scoring algorithms. Specifically, average the responses for questions Q5-Q8, Q10-Q13, and Q15-17 (0=never, 25=rarely, 50=sometimes, 75=often, 100=always). If there are at least 9 non-missing responses, then the score is the average of the non-missing responses, otherwise, the score is not calculated. The outcome will then be computed as the difference in score at each post-baseline visit through visit 15 and the score at baseline. If data at any time point are missing, the outcome variable will be coded as missing.

Variable	Type	Definition
Use of Supplementary Treatments for ABL through Visit 15	Dichotomous	The use of supplemental ABL treatments will be computed at baseline and at each visit during which assessments are administered and collected. Specifically, the responses to QA2b from the St. Mark's (Vaizey) questionnaire will directly provide this outcome. If data at any time point are missing, the outcome variable will be coded as missing.
Patient Global Impression of Improvement through Visit 15	Ordinal	The outcome will be the reported PGI-I value at each visit during which assessments are administered and collected. If data at any time point are missing, the outcome variable will be coded as missing.
Patient Global Impression of Improvement through Visit 15	Dichotomous	For defining improvement, Yes="much better" or "very much better" on the PGI-I scale. The outcome will be computed at each visit during which assessments are administered and collected. If data at any time point are missing, the outcome variable will be coded as missing.
Patient Global Symptom Control through Visit 15	Ordinal	The outcome will be the reported PGSC value at each visit during which assessments are administered and collected. If data at any time point are missing, the outcome variable will be coded as missing.
Change from baseline in SF-12 PCS Subscale through Visit 15	Continuous	*See Attachment 2.*
Change from baseline in SF-12 MCS Subscale through Visit 15	Continuous	*See Attachment 2.*
Treatment adherence through Visit 15	Dichotomous	Adherence to treatment through Visit 15 will be defined as attending at least 10 of 12 treatment sessions in Part I.

Variable	Type	Definition
Treatment adherence through 12 Months	Dichotomous	Adherence to treatment through 12 months will be defined as attending at least 9 of 11 treatment sessions in Part II (Fixed schedule only)
Analysis of masking procedure	Categorical	Percent of participants who responded that they thought they were assigned the PTNS or sham treatment or did not know which treatment had been assigned. (correctly guessed)
Change in St. Marks and Diary Outcomes through Visit 15	Continuous	The change from baseline outcomes calculated through Visit 15 will be extended to be calculated through 12 months
Rate of UTIs treated with antibiotics	Categorical	Percent of participants reporting a UTI treated by antibiotics on the Medical Follow-up Form
Cost of PTNS Maintenance Schedules	Continuous	This is calculated as the number of PTNS treatments through 12 months for Part II.

9.3 Primary Analysis Methods

For the primary aim in Part I, the change from baseline in St. Mark's (Vaizey) score after 12 weeks of stimulation will be compared between the PTNS and sham groups using a longitudinal general linear model. For the purposes of this analysis, time will be treated as both a categorical repeated-measure as well as a continuous term in order to account for correlation between measurements and allow for modeling of trajectory with respect to time. The potential quadratic effect with respect to time will be assessed and will be retained in the final model if a quadratic effect is evident. The model will include fixed effects for treatment group, time, and interactions between those variables. Thus, the models will allow for different trajectories of change for women who are or are not randomized to PTNS. The model will also be adjusted for the design effect of stratification by center. A statistical test based on the model will be conducted to assess whether mean changes from baseline in St. Mark's (Vaizey) scores at 12 weeks are significantly different between the two treatment groups. A sensitivity analysis will be conducted in which additional terms will be included in the model for type of Run-In Phase diary (paper or phone app), and interactions between time, treatment group, and type of Run-In Phase diary. If there are important baseline characteristics that differ between randomized treatment groups by chance, additional sensitivity analyses may be conducted to evaluate the impact of those imbalances.

For Part II, the percent of responders and a 95% confidence interval will be estimated in each PTNS maintenance group at one year using Wilson score intervals. Participants who were not treated with PTNS will be excluded from Part II analyses. Participants who begin but do not complete Part II will be considered non-responders for analysis purposes; however, sensitivity analyses will be conducted to assess the robustness of the Part II results to this assumption.

For Part III, the association between duration of treatment effect and patient characteristics will be explored. Participants who were not treated with PTNS and who were not still responding to treatment at the end of Part II will be excluded from Part III analyses. Participants who begin but do not complete Part III will be considered to fail one day after their last collected time point for analysis purposes; however, sensitivity analyses will be conducted to assess the robustness of the Part III results to this assumption.

For the Run-In phase, change in St. Mark's score and weekly FIEs from week 1 to week 4 will be estimated using general linear modeling. To assess differences between the modalities of the bowel diary (paper or phone app), type of diary will be included in the general linear models as an independent variable along with characteristics that differ between the groups.

Pearson correlation will be used to determine if the changes from baseline in FIE recorded on the PFDN ABL app diary correlate with changes from baseline in other measures of FI symptom severity

9.4 Secondary Analysis Methods

For Part I, changes from baseline in secondary outcomes after 12 weeks of stimulation will be compared between treatment groups using models similar to the primary outcome for continuous measures and analogous generalized linear models for categorical outcomes. Analyses of secondary outcomes will be considered exploratory, and confidence intervals and p-values will be presented for descriptive purposes.

For Part II, models analogous to the ones described for Part I will be used to estimate changes from baseline in other functional outcomes at one year for each PTNS maintenance group. Time points through one year will be included in analysis, and models will include terms for maintenance group, time, and the interaction between maintenance group and time.

For Part III, end of effect will be defined as the first time point at which a subject reports a PGSC ≤ 2 or reports initiating a NEW treatment for ABL that is prescribed by a health provider. Association between loss of symptom control and length of time since last PTNS session will be modeled using a generalized linear mixed model. The model will include all time points assessed after the end of treatment and will account for correlations between repeated measurements on the same subject. Continuous outcomes such as change in St. Mark's score will be modeled using analogous general linear mixed models.

10 SAFETY ANALYSES

10.1 Overview of Safety Analysis Methods

All safety analyses will be performed using all participants who were randomized, regardless of eligibility. Descriptive p-values comparing the study arms will be provided on most safety table summaries and will be obtained using chi-square tests for binary outcomes. If the number of events allow, a 2-sided Cochran Mantel-Haenszel test controlling for strata defined and site will be used to obtain the p-values.

10.2 Adverse Events

Per the protocol, participants were asked to report any adverse events from initiation of treatment through 12 months follow-up. All adverse events were collected on an adverse event log and coded using MedDRA (V17.0).

AEs will be listed and summarized by system organ class and preferred event term. Summaries will be of the number of events and number of individuals experiencing events by treatment group and will be created for all AEs, AEs by severity, and AEs by relationship to treatment. Any events starting outside of the reportable time frame will be included in separate listings and will be excluded from summary tables. If a complete onset date is unknown and it cannot be confirmed that the event occurred during this time period, then the event will be considered a treatment-emergent AE.

10.3 Deaths and Serious Adverse Events

A serious adverse event (SAE) is any event that is life threatening, results in death, causes or prolongs hospitalization, leads to a disability or birth defect, or requires an intervention to prevent a disability. SAEs will be listed and SAEs and treatment-related SAEs will be summarized in the manner mentioned in Section 10.2 if there are enough events to summarize. Deaths will be listed.

11 PHARMACOKINETIC ANALYSES

No pharmacokinetic analyses are planned.

12 ANALYSIS OF OTHER OUTCOMES

No analyses of outcomes other than efficacy and safety outcomes are planned.

13 REPORTING CONVENTIONS

Unless required otherwise by a journal, the following rules are standard:

- Moment statistics including mean and standard deviation will be reported at 1 more significant digit than the precision of the data.
- Order statistics including median, min and max will be reported to the same level of precision as the original observations. If any values are calculated out to have more significant digits, then the value should be rounded so that it is the same level of precision as the original data.
- Following SAS rules, the median will be reported as the average of the two middle numbers if the dataset contains even numbers.
- Test statistics including t and z test statistics will be reported to two decimal places.
- P-value will be reported to 3 decimal places if > 0.001. If it is less than 0.001 then report '<0.001'. Report p-values as 0.05 rather than .05.
- No preliminary rounding should be performed, rounding should only occur after analysis. To round, consider digit to right of last significant digit: if < 5 round down, if >=5 round up.

14 CHANGES TO THE ANALYSIS PLANNED IN THE PROTOCOL

An additional analysis will be performed wherein minimal important difference (MID) will be calculated for diary metrics such as the number of fecal incontinence episodes per week and the number of urgency episodes per week. The MID will be determined using either anchor-based or distribution-based methods, or both.

15 REFERENCES

None.

16 LIST OF POTENTIAL DISPLAYS

Data displays may be added, deleted, rearranged or the structure may be modified after finalization of the SAP. Such changes require no amendment to the SAP as long as the change does not contradict the text of the SAP.

Tables
Participant Eligibility
Participant Disposition
Protocol Deviations
Demographic and Baseline Characteristics
Primary Efficacy Model Results
Secondary Efficacy Outcome Measures
Safety Summary
Adverse Events
Figures

Consort diagram of participant disposition
Data Listings
Subject Eligibility
Subject Disposition
Protocol Deviations
Adverse Events
Serious Adverse Events

17 ATTACHMENTS

17.1 Attachment 1: Outcome Calculation for Diary Outcomes

The NOTABLe diary mobile application provides a new and improved approach in collecting incontinence events from participants in real time. Rather than relying on paper diaries, the application provides participants the ability to log events on their phone.

This was the first application of a diary collection via mobile device for the PFDN. The data from each individual's diaries is uploaded to a central database and pushed to the Medidata EDC. Due to some improvements in the data upload partway through the study and to how data was recorded when subject used extended app button presses or quick, repeated presses of app buttons, the uploaded data may contain gaps in time. These gaps could represent real, missed days of diary collection as well as days in which a participant indicated that no event occurred. The distinction between these two types of events is important in determining average daily and weekly events as well as overall compliance.

In order to ensure data integrity, a manual adjudication process will be performed for each diary collected as part of Part I and Part II of the study. The adjudications will be made independently by three DCC study staff (study coordinator, study statistician, and junior study statistician) and then collectively reviewed to ensure consensus. To avoid bias, adjudication will not consider treatment group in determining app data behavior and the adjudication process will generally follow the rules listed below. Special cases with many gaps or unexpected data behavior will be reviewed jointly by the committee.

These are the general adjudication rules:

1. Null entries (rows that could indicate skipped days or days with no event) will be set to the first date prior to the null entry. Prior to January 2019, these rows are identified by a row of 0 values. After January 2019, these are identified by a 99 for the Bristol Stool value.
2. If the first record for a diary is null, the date will be set to the date prior to the null entry in consultation with the diary configuration file which indicates start and stop days for each diary.
3. If multiple null entries follow a gap in time of records, the first null entry will be set to the day before the null entry and all other null entries will be removed.
4. Zero rows after the app update implementation date January 2019 will be removed and the temporally adjacent (within a few minutes) record immediately following the zero row will be set to the date before. Zero rows without temporally adjacent records will be removed.

17.2 Attachment 2: Scoring of the SF-12 PCS and MCS Subscales

The scoring of the SF-12 QOL assessment was found at https://drhays.dgsom.ucla.edu/files/view/docs/programs-utilities/sf12v1_short.sas.txt. This attachment provides a summary of the calculations for both the Physical Component Summary (PCS) and Mental Component Summary (MCS) subscales of the SF-12.

The SF-12 is comprised of 12 questions. Q1, Q8, and Q12 use five-number scales; Q4 through Q7 use yes/no scales; and Q 2 and Q3 have three-number scales; and Q9 through Q11 use six-number scales. For Q1 and Q8, values are assigned to each response as follows: 100=excellent, 75=very good, 50=good, 25=fair, 0=poor. For Q12, values are assigned to each response as follows: 100=none of the time, 75=a little of the time, 50=some of the time, 25=most of the time, 0=all of the time. Q2 and Q3 receive values as follows: 0=yes, limited a lot; 50=yes, limited a little; 100=no, not limited at all. For Q4-Q7, values are assigned as follows: 0=yes and 100=no. For Q9 and Q10, values are assigned to each response as follows: 100=all of the time, 80=most of the time, 60=a good bit of the time, 40=some of the time, 20=a little of the time, 0=none of the time. For Q11, values are assigned to each response as follows: 0=all of the time, 20=most of the time, 40=a good bit of the time, 60=some of the time, 80=a little of the time, 100=none of the time.

With values assigned to each question, the raw PCS and MCS scores are then calculated as a weighted sum based on the assigned value for each question. The weights are provided in Table A-1 and Table A-2. The PCS and MCS subscales are then calculated by multiplying these sums by 10 and adding a constant to each raw score (PCS=56.57706, MCS=60.75781).

Table A-1. Sum Weights for Question Values – PCS

Question Number	Assigned Value						
	0	20-25	40	50	60	75-80	100
1	-8.37399	-5.56461		-3.02396		-1.31872	0
2	-7.23216			-3.45555			0
3	-6.24397			-2.73557			0
4	-4.61617						0
5	-5.51747						0
6	3.04365						0
7	2.32091						0
8	-11.25544	-8.38063		-6.50522		-3.80130	0
9	3.46638	2.90426	2.37241		1.36689	0.66514	0
10	-2.44706	-2.02168	-1.61850		-1.14387	-0.42251	0
11	4.61446	3.41593	2.34247		1.28044	0.41188	0
12	-0.33682	-0.94342		-0.18043		0.11038	0

Table A-2. Sum Weights for Question Values – MCS

Question Number	Assigned Value						
	0	20-25	40	50	60	75-80	100

1	-1.71175	-0.16891		0.03482		-0.06064	0
2	3.93115			1.86840			0
3	2.68282			1.43103			0
4	1.44060						0
5	1.66968						0
6	-6.82672						0
7	-5.69921						0
8	1.48619	1.76691		1.49384		0.90384	0
9	-10.19085	-7.92717	-6.31121		-4.09842	-1.94949	0
10	-6.02409	-4.88962	-3.29805		-1.65178	-0.92057	0
11	-16.15395	-10.77911	-8.09914		-4.59055	-1.95934	0
12	-6.29724	-8.26066		-5.63286		-3.13896	0