STATISTICAL ANALYSIS PLAN

Addendum 1 (V 1 12SEP2015)

PFDN Protocol Number 22P01:

Refractory Overactive Bladder: Sacral NEuromodulation v. BoTulinum Toxin Assessment (ROSETTA)

AUTHOR (S): Tracy Nolen
Dennis Wallace

1 MODIFICATIONS AND CLARIFICATIONS TO PLANNED ANALYSES

1.1 Bladder Diaries

Two primary types of activities have occurred as part of QA/QC processes for the ROSETTA bladder diary data that have impacted how we will analyze the study data. The first set of activities include a central audit of bladder diaries and the second being a set of visits to the Kaiser-Downey (Bellflower) site.

During the centralized audit of bladder diaries, it was identified that:

- Source diaries could not be identified for participant 25RS047 at 4 and 5 months.
- Participant 07RS054 only had 2 valid days of incontinence episode data reported at baseline.

The first site visit to Kaiser was conducted October 6-7, 2014. During the visit a number of potential issues regarding data quality at the site were identified. Based on these concerns, a second, for cause, site visit was scheduled to the Bellflower facility during November 2014. During that site visit, all available study participant diaries from all randomized ESTEEM and ROSETTA participants were reviewed.

Key findings that impact analysis approach decision for ROSETTA included:

- Source diaries could not be found for 23 follow-up and 7 baseline visits. After the visit, a subset of these diaries were located and determined to be valid diaries. The final set of missing diaries include:
    - 20RS001: 9 months
    - 20RS007: 2-12 months
    - 20RS009: 2 and 12 months
    - 20RS011: 3 months
    - 20RS014: 1 month
    - 20RS015: 1 month
    - 20RS018: Baseline
    - 20RS019: Baseline
    - 20RS024: Baseline
    - 20RS025: Baseline
- Follow-up for all Interstim participants were completed using 7 day diaries instead of 3 day diaries, a clear protocol deviation that can bias study results
- Participant 20RS023 was determined to be ineligible at baseline based on corrections made to extracting data from the baseline diary.

Based on these findings the following decisions regarding analyses were made:

- Data without a source diary will be excluded from the primary bladder diary analyses. Participants with missing baseline diaries will be completely excluded from these analyses as change from baseline outcomes cannot be calculated reliably at any visit. Sensitivity analyses that summarize the change in summary statistics for the bladder diary endpoints when

including and excluding these diaries will be done to assess the impact of the exclusion of these diaries on study results.

- Participants identified during the audit or site visits as being ineligible at baseline based on a review of diaries will be excluded from all study analyses of efficacy. If treated, they will remain in the safety population.
- Baseline diaries found during the audit or site visits to only have two valid days for randomized participants will be considered valid for the purposes of identifying eligibility as well as for use in study analyses.

1.2 Masked Data Review Findings

After the database was locked for the 6 month primary outcome analyses but prior to analysis of any study data, the protocol team reviewed key data in a masked fashion (e.g. with dummy IDs and no information about treatment assignment where possible) to make final decisions regarding analysis population membership as well as to address any outstanding questions with respect to which data should be included in the analyses. The final decisions are below:

1.2.1 Analysis Population Membership:

Safety, ITT and Clinical Responder Populations:

- Per the SAP, individuals not receiving full treatment on Interstim are excluded from the CR population. Although not mentioned in the SAP, we are similarly excluding Botox individuals that did not receive a complete 1st injection.
- In reviewing the protocol deviations for the PP population, a deviation was noted that the reported response diary for a participant was invalid. Participant 24RS045 will be classified as a non-responder and be excluded from the CR population.

PP Population:

- UI Therapies: A DCC nurse coordinator reviewed all reported therapies to ensure that all urinary incontinence (UI) therapies were appropriately identified. The study PI reviewed all UI therapies.
  - Any data collected after the use of the following medications will be excluded from PP analyses (anticholinergics and beta-3 adrenergic agonist specifically will be summarized for the primary manuscript and are identified in red font below):

200 UNITS BOTOX INJECTION	INTERSTIM IMPLANTATION.	OXYBUTININ PATCH
2ND INTERSTIM PLACEMENT	INTERSTIM PLACEMENT	OXYBUTYNIN
BOTOX	INTERSTIM PLACEMENT (STAGE I AND II)	OXYBUTYNIN (ONLY BEFORE STUDY TREATMENT AND DIARY)
BOTOX INJECTION	INTERSTIM STAGE 1 IMPLANTED	OXYBUTYNIN(DITROPAN)
BOTOX INJECTION (IN BLADDER)	INTERSTIM STAGE 2 IMPLANTATION	OXYTROL PATCH
BOTOX INJECTION 200U	INTERSTIM STAGE 2 PLACEMENT	PERCUTANEOUS TIBIAL NERVE STIMULATION (PTNS).
CROSSOVER TO INTERSTIM DEVICE	INTERSTIM STAGE 2 SURGERY	PTNS SESSIONS
DARIFENACIN (ENABLEX)	MIRABEGRON	PTNS SESSIONS.
DETROL	MIRABEGRON 25MG 1 PO EVERYDAY	SANCTURA

DETROL LA	MIRABEGRON 50 MG 1 DAILY	SANCTURA 1 PO BID
DETROPAN XL	MIRABEGRON E.R.	SOLFENACIN (VESICARE)
DITROPAN	MIRABETRIQ- 25MG	SOLIFENACIN (VESICARE).
DITROPAN XL.	MIRABEYRON	TOLTERODINE
ENABLEX	MIREBEGRON	TOLTERODINE (DETROL LA)
FSLP	MYBETRIQ	TOVIAC
FSLP/INTERSTIM	MYBETRIQ 50 MG 1 EVERY OTHER DAY	TOVIAZ
GELNIQUE	MYRBETRIQ	TOVIAZ 8MG PO ER 1 PO DAILY
GELNIQUE, APPLY 1 PK OF GEL TO SKIN DAILY	MYRBETRIQ - 50MG	TOVIAZ 8MG PO QD
INTERSTIM	MYRBETRIQ -25MG	TROPIUM (SANCTURA)
INTERSTIM 1	MYRBETRIQ 25MG	TROSPIUM
INTERSTIM 2	ONABOTULINUMTOXINA (BOTOX)	TROSPIUM (SANCTURA XR)
INTERSTIM BATTERY AND LEAD WIRE REPLACED	OXYBURIN GHS	TROSPIUM (SANCTURA)
INTERSTIM DEVICE	OXYBUTANIN	TROSPIUM (SANCTURA).
INTERSTIM DEVICE STAGE I	OXYBUTIN	VESICARE

- Although reported as UI therapies, data will not be excluded after the use of the following treatments:

ACCUPUNCTURE
CAPSAICIN .025 % EX CREAM
COAPTITE INJECTION
DESIPRAMINE (NORPRAMIN)
ESTRONG VAGINAL RING
IMIPRAMINE
KEGELS (SELF-ADMINSTERED)
MID-URETHRAL SLING
NORTRIPTYLINE
PELVIC FLOOR PHYSICAL THERAPY
PESSARY
PREMARIN

- Interstim Device Usage: Visits with the Interstim device turned on <50% of the time since last visit are to be excluded from PP analyses. There are some visits where device use was not collected.

- Visits where device use information was not available due to a missed visit or programmer not being brought in will use device use % from the next attended visit if one exists. Otherwise device use will be assumed to be 0%. For data through 6 months on participants not excluding from the per-protocol population for other reasons:
    - Missed visits or visits with no programmer with no subsequent completed visit: 25RS032: Months 3-6; 26RS024: 3-6.
    - Missed visits or visits with no programmer with subsequent completed visit: 02RS052- Months 1-3 use; 02RS063- Months 3-6; 02RS066- Months 1-3; 07RS044- Months 0-1; 08RS014- Months 0-1; 15RS016- Months 1-3; 15RS024- Months 1-6; 20RS009- Months 0-1; 23RS051- Months 3-6; 24RS033- Months 1-3; 24RS036- Months 1-3; 24RS041- Months 3-6 (visit completed but participant did not have time to do programming assessment but site noted everything working fine); 25RS021- Months 0-1 (visit completed but participant did not have time to do programming assessment but site noted everything working fine); 25RS032- Months- 0-1; 25RS048- Months 1-3.
- Visits with no documented reason for why programming information was not reported will be assumed to be 0%. For data through 6 months on participants not excluding from the per-protocol population for other reasons:
    - 02RS050- Months 0-1 use; 07RS031- Months 1-3; 24RS030- Months 0-1.
- Visits with no programming information due to technical difficulty will be assumed to be 0% as we are unsure if device was reset. For data through 6 months on participants not excluding from the per-protocol population for other reasons:
    - 08RS008- Months 0-1 use (device would not sync)
    - 14RS012- Months 0-1 use (programmer device broken)
    - 25RS050- Months 3-6 use (malfunction of programmer)
    - 25RS052- Months 1-6 use (programmer not brought to 3 month and then malfunctioned at 6 month)
    - 25RS053- Months 3-6 use (malfunction of device)
    - 25RS054- Months 0-3 use (programmer not brought to 1 month visit and then printer didn't work at 3 month)

- Protocol Deviations: The study PI reviewed all eligibility, randomization and treatment administration deviations as well as any deviations that mentioned key words (e.g. Interstim, Botox, injection, implantation) to identify any additional deviations that would excluded all or part of a participant's data from the PP analyses. Likewise, the DCC study team reviewed all other deviations. The following decisions were made:
    - All data after a participant receives an incomplete or non-protocol dose re-injection will be excluded from PP analyses (e.g. Subject 25RS009).
    - Study therapy deviations related to safety or operational requirements that should not impact the therapy efficacy will not have data excluded as a result (e.g. botox procedure

in an individual that was culture positive, Subject 08RS003 (excluded for other reason: incomplete injection); <2 programs determined during FSLP test period, Subject 08RS011; different gauge needle or dilution amount used than that specified in protocol, subject 16RS014 and 24RS015).

1.2.2 Data Inclusion Decisions: Diary

- Participants with delays in treatment (or decision to not treat) more than 90 days from randomization were to have their baseline diary repeated. Diaries were not repeated for 27 participants. Among those ultimately treated, all participants were treated within 180 days of completing their baseline diary except for one individual who was treated at 302 days post-diary completion. The protocol team reviewed lags between baseline diaries and treatment initiation and determined that all baseline diaries would be included in the analysis and no participants would be excluded from any analysis population based on the timing of their baseline diary.
- Diaries not completed on actual forms will be included in analyses.
- Although the MOP specified follow-up diaries must have 2 **consecutive** days to be considered valid, any recorded diaries with valid data for at least 2 days regardless of if consecutive will be analyzed.
- As detailed in the SAP, outcome data are analyzed based actual study month within which they were obtained as opposed to reported study month. If there are no outcome data within a visit window, then data outside the visit window may be used. Clarifications to the analysis plan:
    - Out of window outcome data will only be used for a particular visit if it does NOT fall into the window of another visit.
    - Visit windows were +/- 10 days through 6 months and +/-28 days. For analysis, the following rules will be employed regarding visit windows:
        - Data more than 5 days outside of the window through 6 month will not be considered for analysis. This implies an overall +/-15 day window through 6 months and excludes very little data as most windows are nearly back to back.
        - Data more than 14 days outside of the window for visits after 6 months will not be considered for analysis implying an overall +/- 42 day window.
- Site specific plots of mean and median diary data outcomes were reviewed (e.g. site median of mean UUIE/day at each follow-up visit). The following observations were explored. No additional items were identified for follow-up.
    - Stress Episodes: Overall the mean numbers are low which suggests we did identify our population of interest. No site trends appear concerning.
    - Urge Episodes: One site has higher visit means compared to the other sites; however, the site's medians do not differ from other sites. There do appear to be more subjects that continue to have high values post-baseline at this site than other sites; however, a single extreme subject seems to be the primary contributor to the site means differing.
    - Other Episodes: One site has higher visit means compared to the other sites; however, the site's medians do not differ from other sites. A single extreme subject seems to be the primary contributor to the site means differing.
    - Day Voids and Pad Use: No site trends appear concerning.

- Night Voids: One site has higher visit means and medians compared to the other sites. There do appear to be more subjects that continue to have high values post-baseline at this site than other sites.

1.2.3 Data Inclusion Decisions: QOL

- Participants with delays in treatment (or decision to not treat) more than 90 days from randomization had NO requirements for repeating baseline QOL assessments. QOL calls occurred more than 90 days prior to treatment or decision to not treat for 22 participants and OABq-SF was completed more than 90 days prior for 53 participants. Among those ultimately treated, all participants were treated within 180 days of completing their baseline QOL call except for two individuals who were treated at 234 and 269 days post-call completion. Among those ultimately treated, all participants were treated within 180 days of completing their baseline OABq-SF except for 4 individuals who were treated at 204, 221, 283, and 288 days post-OABq-SF completion. The protocol team reviewed lags between baseline QOL assessments and treatment initiation and determined that all baseline QOL assessments would be included in the analysis and no participants would be excluded from any analysis population based on the timing of their baseline QOL assessment.
- QOLs filled out with assistance or on paper will be included in all analyses.
- As detailed in the SAP, outcome data are analyzed based actual study month within which they were obtained as opposed to reported study month. If there are no outcome data within a visit window, then data outside the visit window may be used. Clarifications to the analysis plan:
  - Out of window outcome data will only be used for a particular visit if it does NOT fall into the window of another visit.
  - Visit windows were +/- 10 days through 6 months and +/-28 days. For analysis, the following rules will be employed regarding visit windows:
    - For QOL calls occurring no more frequently than 6 months apart:
      - Data more than 32 days outside of the window at 6 month will not be considered for analysis. This implies an overall +/-42 day window at 6 months
      - Data more than 14 days outside of the window for visits after 6 months will not be considered for analysis implying an overall +/- 42 day window.
    - For OABq-SF, the same rules defined for diaries above will be employed as this assessment was done on the same time table as diaries.
- Site specific plots of mean and median QOL data outcomes were reviewed. No site trends have been identified of concern. No additional items are identified for follow-up.

1.2.4 Data Inclusion Decisions: Other Points

- Incomplete or invalid UDS at baseline will not preclude an individual from analyses except for those that use UDS as a potential predictor.
- Per protocol and MOP, the definition of a UTI event of interest is symptomatic, culture + (or culture could not be performed) and treated. Since sites generally took a broader reporting approach, we have reviewed and marked whether each event meets the protocol definition.

Detailed notes are in the UTIS (MaskedReview_UTIs_wMaskedNotes.docx) but generally, if the protocol definition could not be confirmed as met (e.g. inconclusive culture or not symptoms specified) then the event is not counted as a UTI.

- AEs of pain, shortness of breath or muscle weakness
    - AEs that may possibly represent pain, shortness of breath or muscle weakness were identified as being of interest for primary manuscript. These events will be summarized by SOC and Preferred Term to determine appropriate classification for manuscript.

1.3 QOL Scoring Updates

1.3.1 OAB-SATq

The scoring manual for the OAB-SATq states that higher scores should correspond to better outcomes and also details that final composite scores for each dimension should be obtained by subtracting raw scores from the highest possible score and dividing by range. However, based on the direction of score for individual responses, this derivation would result in lower scores corresponding with better outcomes. We have adjusting the scoring algorithm instead to be: 100*(raw score – lowest possible score)/score range. This calculation results in higher scores corresponding to better outcomes and an overall scale of 0-100.