Bio-Medical Research Ltd

STATISTICAL ANALYSIS PLAN

Version 1.0

22-Jan-2016

Investigational Product:	Vital Compact
Protocol Number:	BMR-13-1001
Protocol Title:	A single-blind, multi-centre, randomised, controlled, non-inferiority, clinical study to assess the safety and performance of the Neurotech Vital Compact device compared to the itouch Sure Pelvic Floor Exerciser for the treatment of stress urinary incontinence in female patients
Protocol Issue Date:	27-Mar-2015
Protocol Version:	2.0

Bio-Medical Research Ltd

Study No: BMR-13-1001

A single-blind, multi-centre, randomised, controlled, non-inferiority, clinical study to assess the safety and performance of the Neurotech Vital Compact device compared to the itouch Sure Pelvic Floor Exerciser for the treatment of stress urinary incontinence in female patients

APPROVAL SIGNATURES

This Statistical Analysis Plan has been reviewed and approved by the following personnel:

Prosoft Personnel

Title	Printed Name	Signature	Date
Senior Director, Prosoft Clinical	Diane Tipping	Diane Tipping	22-Jan-2016

Bio-Medical Research Personnel

Title	Printed Name	Signature	Date
Managing Director, Expedite Research	Barbara Powers	Barbara Powers	25 JAN 2016

Title	Printed Name	Signature	Date
Technical Director, Bio-Medical Research Ltd	Conor Minogue		8 Feb 2016

1 Table of Contents

1 INTRODUCTION

1.1 STUDY BACKGROUND

Stress urinary incontinence often leads to subjects limiting their participation in social activities, which in turn has a subsequent impact on a person's quality of life. While some seek surgical intervention, physical therapy is seen as the first treatment option in this often under-reported condition. The purpose of this study is to conduct a pivotal study comparing the Vital Compact device (fourth generation device) to an FDA 510(k) approved predicate device (itouch Sure Pelvic Floor Exerciser).

1.2 STUDY OBJECTIVES

The primary objective of this clinical study is to evaluate and compare the safety and performance profile of the Vital Compact device (delivering electrical stimulation through external electrodes) and the itouch Sure Pelvic Floor Exerciser (delivering electrical stimulation through an internal vaginal probe) for the treatment of stress urinary incontinence following a 12-week treatment program.

1.3 STUDY DESIGN

This is a prospective, randomized, controlled, single-blind, multi-site clinical study to be conducted in the United States of America (USA) employing Neuromuscular Electrical Stimulation (NMES) to stimulate the pelvic floor muscles of women suffering from stress urinary incontinence.

Approximately one-hundred and eighty (180) female subjects diagnosed with stress urinary incontinence will be enrolled in this study. All subjects who are considered eligible to participate in the clinical study and give consent will be randomised to complete either a 12-week treatment programme with the Vital Compact device or a 12-week treatment programme with the itouch Sure Pelvic Floor Exerciser. The 12-week treatment programme will be completed by the subjects at home with treatment with the device in accordance with the device Instructions for Use. Subjects randomized to the Vital Compact device will be instructed to use the device at home once per day, in a therapeutic position, for five days each week with two rest days taken within a 7 day week period, for the 12-week period; each treatment cycle is fixed at 30 minutes and the device will alert the subject when the cycle is completed. Subjects randomized to the itouch Sure Pelvic Floor Exerciser will be instructed to use the device at home once per day for the 12-week period; each treatment cycle is fixed at 20 minutes and the device will alert the subject when the cycle is completed.

Subjects included in the clinical study will be evaluated at screening, on enrolment into the study (baseline) and during the 12-week treatment programme at 4 and 12 weeks. A telephone call will be made at 1 week to check on the subject's progress. In addition, subjects will be evaluated at 26 weeks following their commencement of the treatment.

1.4 SAMPLE SIZE AND POWER

Sample size calculations were performed using the NQuery Advisor® version 5.0, under a two group test of proportions.

Research suggests that the response rates for Vital Compact and itouch Sure at 12 weeks will be 71% and 46%, respectively (itouch success rate based on Sand, 1995).

The sample size calculation is based on the assumption that the itouch Sure Pelvic Floor Exerciser success rate is 52% and the Vital Control success rate is 71%. A sample size of 87 subjects per group will provide 90% power using a one-sided Type I error rate of 0.025, and a non-inferiority margin of 5%. A randomization allocation of 1:1 was also assumed. For practical reasons, the sample size has been increased to a recruitment of 180 subjects (90 subjects per treatment group).

1.5 STUDY POPULATION

1.5.1 MAIN INCLUSION CRITERIA

Female subjects between the ages of 18 and 65 who have been clinically diagnosed with stress urinary incontinence and have failed to improve their condition using Kegel exercises.

1.5.2 MAIN EXCLUSION CRITERIA

Subjects with medical, physical, or neurological conditions that would compromise their participation or make them unable to perform the study procedures.

1.6 STATISTICAL ANALYSIS PLAN (SAP)

1.6.1 SAP OBJECTIVES

The objective of this SAP is to describe the statistical analyses to be carried out under BMR-13-1001.

1.6.2 GENERAL PRINCIPLES

Data will be summarized using descriptive statistics. For continuous measures, this will include number (N), mean, standard deviation, median, and range (minimum and maximum). For categorical measures, counts and percentages will be provided. Data will be summarized separately for each randomized arm under the principles of intent to treat, unless otherwise noted. All statistical tests will be conducted at $\alpha = 0.05$, two-sided significance level, unless otherwise stated. The day of randomization is considered as Day 1 and the day prior to randomization is Day -1. Summaries will be based on the database visits, regardless of the relative day of the visit; early termination visits are to be collected as Week 12 and all follow-up assessments are collected as Week 26.

1.6.3 CURRENT PROTOCOL

The current study protocol at the time of writing this SAP is version 2.0 FINAL 27MAR2015. Any future amendments to the protocol will be reviewed for their impact on this SAP, which will be updated only if necessary.

1.6.4 SOFTWARE

Analyses will be conducted using SAS for Windows version 9.3 or later. In the event that an analysis is required that is better suited for a statistical package other than SAS, another package (e.g. R) will be used.

2 ANALYSIS

2.1 ANALYSIS POPULATIONS

The primary analysis population will be the intent to treat (ITT) population. This will include all randomized subjects. Subjects will be summarized according to their randomized assignment. If there are randomized subjects who do not use the device at least once, a modified ITT (mITT) population will be defined to exclude such subjects and this population will be used in place of ITT.

A per-protocol (PP) population will also be evaluated for the primary endpoint. This population will consist of subjects who:

- Signed informed consent
- Met inclusion/exclusion criteria
- Were treated per randomized assignment
- Had 12 week outcome data available
- Met study compliance criteria
- No significant protocol deviations

Prior to locking the database, the PP criteria will be reviewed, additional criteria will be added if needed, and a listing of PP violations will be generated. The Sponsor/CRO will review this list and determine exclusions from this population. The final list of PP criteria and exclusions will be documented prior to unblinding.

The Safety population will include all randomized subjects who used the device at least once. Subjects will be summarized according to the actual device used, rather than the randomized assignment.

2.2 SUBJECT ACCOUNTING AND BASELINE CHARACTERISTICS

The ITT and PP populations will be summarized for the following demographic and baseline characteristics:

- age
- race and ethnicity

- menopause status
- height, weight, and BMI
- general medical history (ITT population only)
- physical examination findings (ITT population only)
- prior and concomitant medications (ITT population only)

Baseline assessments of stress urinary incontinence data will be summarized for the ITT and PP populations. Incontinence Quality of Life Questionnaire (I-QOL) scores and Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire (PISQ-IR) scores will be summarized for the ITT and PP populations. Study completion status and primary reason for discontinuation will be summarized for the ITT population. Medical history will be coded using Medical Dictionary for Regulatory Activities (MedDRA) and medications will be coded using the World Health Organization Drug dictionary (WHO Drug).

2.3 EFFICACY ENDPOINTS

2.3.1 PRIMARY ENDPOINT

The primary endpoint is defined as the proportion of subjects considered to have achieved 'significant improvement' following a provocative pad weight test at 12 weeks compared to baseline. 'Significant improvement' is defined as greater than 50% reduction in urine leakage from baseline.

2.3.2 PRIMARY ENDPOINT HYPOTHESIS

The primary hypothesis is that the proportion of subjects who respond (i.e. have achieved 'significant improvement') using Vital Compact is not less than 5% worse than the itouch Sure Pelvic Floor Exerciser (i.e. the lower bound of the confidence interval about the difference between device groups should be greater than -5%):

$$H_0: p_{Vital\ Compact} - p_{control} \leq -5\%$$

$$H_1: p_{Vital\ Compact} - p_{control} > -5\%$$

where $p_{Vital\ Compact}$ is the proportion of subjects who respond with Vital Compact and $p_{control}$ is the proportion of subjects who respond with itouch Sure Pelvic Floor Exerciser.

If the primary hypothesis of non-inferiority is met, a hierarchically-nested test of superiority will be performed. The hypothesis for the assessment of superiority is:

$$H_0: p_{Vital\ Compact} - p_{control} \leq 0\%$$

$$H_1: p_{Vital\ Compact} - p_{control} > 0\%$$

2.3.2.1 PRIMARY ENDPOINT ANALYSIS

To test these hypotheses, the difference in the proportions of responders and a 95% Confidence Interval (CI) of the difference in proportion of responders will be calculated using the normal approximation to the binomial distribution. Non-inferiority will be claimed if the lower limit of

the 95% CI is greater than -5%. Furthermore, superiority will be claimed if the lower limit of the 95% CI is greater than 0%. The p-value for superiority will also be reported, based on a Chi-Square test. The total number of successes in each study arm will be summarized, along with the corresponding proportions and exact 95% CIs.

The ITT population will serve as the primary analysis population. For this analysis, if a subject does not complete the 12 week assessment, their outcome will be imputed using multiple imputations (Schafer, 1997). The multiple imputation methodology will be implemented as follows:

1. Ten imputed datasets will be generated to impute missing urine leakage using the below SAS code. Only the imputed values at Week 12 will be retained.

```
Proc MI NIMPUTE=10 SEED=20150930 MIN=0 0 0  MAX= . . .;
     var base w4 w12;
     fcs reg(base) reg(w4) reg(w12);
run;
```

2. The ten imputed datasets will be analyzed separately. Using the methodology described above, the Week 12 urine leakage will be compared against the baseline weights to classify subject responders.
3. The proportion of responders in each treatment group, along with the associated standard error, will be calculated for each of the ten imputed datasets. The individual proportion estimates and standard errors will be pooled by treatment using the SAS code below.

```
Proc MIANALYZE;
     modeleffects prop;
     stderr se_prop;
     by trt;
run;
```

4. The difference in the proportions of responders and the standard error of the difference will be calculated for each of the ten imputed datasets. The individual estimates of the difference in proportions, as well as the associated standard errors of the differences, will be pooled using the SAS code below.

```
Proc MIANALYZE;
     modeleffects prop_diff;
     stderr se_diff;
run;
```

A supportive analysis will also be conducted using similar methodology in the PP population.

2.3.2.2 SENSITIVITY ANALYSES

Sensitivity analyses will be performed to assess the impact of missing data on the primary endpoint results. These analyses will be conducted on the ITT population. Two techniques will be employed: the Last Observation Carried Forward (LOCF) method and the Tipping Point analysis (Yan, 2009).

The LOCF method will impute any missing Week 12 urine leakage with the most recent value collected. The difference in the proportions of responders and a 95% CI of the difference in proportion of responders will be calculated using the normal approximation to the binomial distribution. The p-value for superiority will also be reported, based on a Chi-Square test. The total number of successes in each study arm will be summarized, along with the corresponding proportions and exact 95% CIs.

The Tipping Point analysis will be performed directly on the binary endpoint of subject responder (i.e. whether or not the subject achieved 'significant improvement'). It will be implemented as follows:

1. Determine the number of missing values in each treatment arm:
   Let n_{1m} = number of missing values in the Vital Compact group
   Let n_{2m} = number of missing values in the itouch Sure Pelvic Floor Exerciser group
2. Impute the missing data in each group so that all possible combinations of responder/non-responder are considered. Starting at 0 and ending at n_{1m}, the responder rate in the Vital Compact group will be changed by adding 1 responder at a time. Meanwhile, the responder rate in the itouch Sure Pelvic Floor Exerciser group will also be changed by adding 1 responder at a time, starting at 0 and ending at n_{2m}. The difference in the proportions of responders and a 95% CI of the difference will be calculated for each case. The algorithm is as follows:
   a. Let s_1 represent the number of responders within the group of missing Vital Compact data.
      Let s_2 represent the number of responders within the group of missing itouch Sure Pelvic Floor Exerciser group data.
   b. Set $s_1 = 0$
   c. Set $s_2 = 0$
   d. Calculate the difference of proportions and associated 95% confidence interval.
   e. $s_2 = s_2 + 1, where\ s_2 \leq n_{2m}$
   f. Repeat steps (d) to (e).
   g. $s_1 = s_1 + 1, where\ s_1 \leq n_{1m}$. Continue to steps (c) to (f).
3. Present the lower 95% confidence limit for each possible missing data combination.

2.3.3 KEY SECONDARY ENDPOINTS

The key secondary endpoints listed below will be tested only if the primary efficacy endpoint is met. In order to control the type I error rate, the endpoints will be analyzed hierarchically. If an endpoint in the sequence fails to show statistical significance, then no statistical claims can be made for the subsequent endpoints in the hierarchy and further testing will be performed for exploratory purposes only. The analyses will be based on ITT population and the LOCF method will be used to impute missing data.

1. Between group comparison of mean change, with respect to baseline, in urine leakage in a provocative pad weight test at Week 12;
2. Within Vital Compact group estimate of mean change, with respect to baseline, in urine leakage in a provocative pad weight test at Week 12;

3. Between group comparison of mean change, with respect to baseline, in urine leakage in the 24-hour pad weight test at Week 12;
4. Within Vital Compact group estimate of mean change, with respect to baseline, in urine leakage in the 24-hour pad weigh test at Week 12;
5. Between group comparison of mean change, with respect to baseline, in the number of incontinence episodes/day recorded using a 7-day voiding diary at Week 12;
6. Within Vital Compact group estimate of mean change, with respect to baseline, in the number of incontinence episodes/day recorded using a 7-day voiding diary at Week 12;
7. Between group comparison of the mean improvement, with respect to baseline, in Incontinence Quality of Life Questionnaire (I-QOL) score at Week 12;
8. Within Vital Compact group estimate of mean improvement, with respect to baseline, in Incontinence Quality of Life Questionnaire (I-QOL) score at Week 12;
9. Between group comparison of mean change, with respect to baseline, in the number of pads used/day recorded using a 7-day voiding diary at Week 12;
10. Within Vital Compact group estimate of mean change, with respect to baseline, in the number of pads used/day recorded using a 7-day voiding diary at Week 12;
11. Between group comparison of the proportion of subjects achieving dryness at Week 12 (<1 g on the provocative pad weight test).

2.3.3.1 Analysis of Between-Group Mean changes for Key Secondary Endpoints

For the between-group comparisons of mean change (endpoints 1, 3, 5, 7, and 9 in the hierarchy noted above), the key secondary hypothesis is that the mean change from baseline to Week 12 for Vital Compact is not equal to that of itouch Sure Pelvic Floor Exerciser:

$$H_0: \mu_{Vital\ Compact} = \mu_{control}$$

$$H_1: \mu_{Vital\ Compact} \neq \mu_{control}$$

where $\mu_{Neurotech\ Vital\ Compact}$ is the mean change from baseline for the Vital Compact group and $\mu_{control}$ is the mean change from baseline for the itouch Sure Pelvic Floor Exerciser group.

To test this hypothesis, a two-sample t-test will be conducted for each endpoint. The difference between device groups will be summarized using mean change, standard deviation, and a 95% confidence interval of the difference. In the event that data are highly non-normal, non-parametric methods (e.g. rank-sum test) may be employed. Normality will be tested using Shapiro Wilk Test.

2.3.3.2 Analysis of Proportions for Key Secondary Endpoints

For the between-group comparison of proportions (endpoint 11 in the hierarchy noted above), the key secondary hypothesis is that the proportion of subjects who achieve dryness using Vital Compact is not less than the proportion who achieve dryness using itouch Sure Pelvic Floor Exerciser:

$$H_0: p_{Neurotech\ Vital\ Compact} - p_{control} \leq 0\%$$

$$H_1: p_{Neurotech\ Vital\ Compact} - p_{control} > 0\%$$

where $p_{Neurotech\ Vital\ Compact}$ is the proportion of subjects who achieve dryness on Vital Compact and $p_{control}$ is the proportion of subjects who achieve dryness on itouch Sure Pelvic Floor Exerciser.

To test this hypothesis, a Chi-Square test will be conducted. The difference between device groups will be summarized and a 95% CI of the difference will be calculated using the normal approximation to the binomial distribution. The counts and proportions of subjects achieving dryness for each device group will be summarized along with the exact 95% confidence interval.

2.3.3.3 Analysis of Within-Group Mean changes for Key Secondary Endpoints

For the within-group comparisons of mean change (endpoints 2, 4, 6, 8, and 10 in the hierarchy noted above), the key secondary hypothesis is that the mean change from baseline to Week 12 within the Vital Compact group is not equal to 0:

$$H_0: \mu_{Vital\ Compact} = 0$$

$$H_1: \mu_{Vital\ Compact} \neq 0$$

where $\mu_{Vital\ Compact}$ is the mean change from baseline for the Vital Compact group.

To test this hypothesis, a one-sample t-test will be conducted for each endpoint. The Week 12 summaries for each device group will include the number of observations, mean, median, standard deviation, minimum, maximum, and 95% confidence interval. In the event that data are highly non-normal, non-parametric methods (e.g. signed-rank test) may be employed. Normality will be tested using Shapiro-Wilk Test.

2.3.4 Other Secondary Endpoints

Additional secondary endpoints will be summarized with descriptive statistics for the ITT population. There are no formal statistical tests associated with these endpoints and missing data will not be imputed.

The secondary endpoints that will be evaluated at 4 weeks and 26 weeks are as follows:

- Proportion of subjects considered to have achieved significant improvement, defined as a greater than 50% reduction in urine leakage from baseline on the provocative pad weight test;
- Urine leakage on the provocative pad weight test;
- Urine leakage experienced by the subject at home in the 24-hour pad weight test;
- Incontinence episodes/day recorded using a 7-day voiding diary;
- Proportion of subjects achieving dryness (<1g urine leakage on the provocative pad weight test);

- Improvement in quality of life assessed using the Incontinence Quality of Life Questionnaire (I-QOL);
- Number of pads/day recorded using a 7-day voiding diary;
- Proportion of subjects achieving dryness, defined as a urine leakage of less than 1.3g on the 24-hour pad weight test;
- Proportion of subjects considered to have achieved a greater than 50% reduction in urine leakage, from baseline on the 24-hour pad weight test.

Additional Week 12 endpoints include:

- Proportion of subjects considered to have achieved a greater than 50% reduction in urine leakage, from baseline on the 24-hour pad weight test;
- Proportion of subjects achieving dryness, defined as a urine leakage of less than 1.3g on the 24-hour pad weight test;
- Proportion of subjects achieving dryness in each group (<1g urine leakage on the provocative pad weight test);
- Mean change, with respect to baseline, in sexual function as measured by the Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire (PISQ-IR);
- Estimates of subject response to the Subject Global Impression of Improvement (PGI-I).

2.4 SAFETY ENDPOINTS

2.4.1 DEVICE USE

Device use will be summarized for each device using the daily device diary for the safety population. The following parameters will be included:

- Average number of sessions per week
- Average duration per session
- Average intensity per session
- Percent of actual use
- Percent of target use

2.4.2 ADVERSE EVENTS

Adverse events will be summarized overall, by severity, seriousness, and by relationship to investigational device. Adverse device effects and early withdrawals due to adverse events will be summarized separately. Ninety-five percent confidence intervals may also be provided for the comparison of serious adverse event rates between groups. Adverse events will be categorized by System Organ Class (SOC) and preferred term using MedDRA. The incidence of adverse events will be summarized overall and by SOC and preferred term using frequencies and percentages. Each subject will be counted only once for each of the incidence rates, regardless

of the number of occurrences (events) the subject experiences. The Safety population will be used.

2.4.3 DEVICE DEFICIENCIES

Inadequacy of medical device with respect to identity, quality, durability, reliability, safety and performance, including malfunctions and inadequate labeling will be presented in a data listing.

2.5 INTERIM ANALYSIS FOR FUTILITY

An interim analysis for futility will be conducted when 90 subjects have Visit 4 (12 week) data available. The analysis will be conducted by an independent statistician. A conditional power will be calculated based on projecting the observed trend of the difference between the two groups in the primary efficacy parameter for subjects completing the study. If the analysis indicates that the conditional power for a positive outcome of non-inferiority is less than 30%, the independent statistician will communicate to a designated sponsor's representative that the study is unlikely to meet its primary objective. The sponsor may evaluate additional information and make a final decision regarding continuation of the study. The interim analysis will not be used to stop the study early for a positive outcome; therefore, no adjustment to the Type-1 error is indicated (Lachin, 2005).

The independent unblinded statistician will receive the interim study data and the randomization information confidentially from the CRO. Using the ITT population with multiple imputations for missing data, the conditional power will be calculated using the normal approximation to the binomial distribution, using the non-inferiority margin of 5% (where the lower limit of the two-sided 95% CI is greater than -5%). As a sensitivity analysis, the conditional power may also be calculated using LOCF for missing data. The independent statistician will communicate the findings directly with the designated sponsor's representative. All communications between the unblinded statistician and the sponsor's representative will be kept confidential and unknown to all personnel involved in the trial (e.g., other sponsor's personnel, CRO, and clinical center personnel).

2.6 POOLABILITY ANALYSIS

A poolability analysis will be performed on the primary endpoint to test for a differential treatment effect across study centres. The primary endpoint will be summarized by study centre. Poolability will be assessed using a logistic regression model. The logistic regression model will include a covariate for the treatment arm, study centre, and interaction effect of treatment by centre. The p-value for the interaction effect will be provided. These analyses will be performed for the ITT population. The LOCF method will be used, imputing missing Week 12 urine leakage with the most recent value collected. A single centre will be used to pool small sites (<4 subjects) for the logistic regression.

2.7 SUBGROUP ANALYSIS

Subgroup analyses will be performed on the primary endpoint ("significant improvement" following a provocative pad weight test at 12 weeks) to test for differential treatment effects across baseline subgroups. Subgroups will be derived for the following parameters: race (white versus non-white), BMI [underweight (< 18.5), normal (18.5 to <25), overweight (25 to <30), obese (≥ 30)], baseline severity of stress urinary incontinence, and average intensity of the stimulation delivered during the 12-week treatment programme [low (≤30 and ≤25 for Vital Compact and itouch Sure, respectively), medium (>30 to ≤60 and>25 to ≤50), high (>60 to ≤90 and >50 to ≤75), and very high (>90 and >75)]. Descriptive statistics will be provided for the primary endpoint for each subgroup. These summaries will be presented for the ITT population using observed data (with no imputation for missing data) and using LOCF to impute missing values.

Logistic regression models will be used to statistically evaluate the effect of the subgroups on the primary endpoint. Specifically, the logistic regression model will include a covariate for the treatment arm, subgroup, and interaction effect of treatment by subgroup. The p-value for the interaction effect will be provided. These analyses will be performed for both the ITT and PP populations. The LOCF method will be used, imputing missing Week 12 urine leakage with the most recent value collected.

The change in urine leakage on the provocative pad weight test at Week 12 will also be summarized in the ITT population for these subgroups. No imputation will be used for missing data.

3 DERIVED VARIABLES

3.1 URINE LEAKAGE AND SIGNIFICANT IMPROVEMENT FROM PROVOCATIVE PAD WEIGHT TEST

The urine leakage from the provocative pad weight test is based on the increase in the weight of pad (including packaging) after the provocative pad weight test. 'Significant improvement' is defined as greater than 50% reduction in urine leakage from baseline.

3.2 URINE LEAKAGE FROM 24-HOUR PAD WEIGHT TEST

The urine leakage from the 24-hour pad weight test is based on the increase in the weight of 8 pads (including packaging) after 24 hours of collection. If fewer than 8 pads are returned, the urine leakage can only be determined if the missing pads were unused; in this case, the return weight is adjusted by the weight of the missing pads (where the average weight of one single dry pad with packaging is 16.7 g).

The 24-hour pad weight test collects urine leakage across three consecutive 24-hour periods at each timepoint. The average of the 24-hour periods at each timepoint will be used for the analysis of this data.

3.3 7-DAY DIARY ENDPOINTS

The number of incontinence episodes (accidental leaks) and the number of pads used in each 24-hour period were to be collected for 7 days prior to Baseline and Weeks 4, 12, and 26. The analyses will be based on the average values from all entries in the 7 day period.

3.4 SEVERITY OF STRESS URINARY INCONTINENCE

The severity of stress urinary incontinence will be defined as dry (< 1.3g), mild (1.3g to < 20g), moderate (20g to < 75g) and severe (≥ 75g), based on the 24-hour pad weight test (O'Sullivan, 2004).

3.5 INCONTINENCE QUALITY OF LIFE QUESTIONNAIRE

The Incontinence Quality of Life Questionnaire (I-QOL) consists of 22 items, all of which are rated on a response scale with five categories (1=extremely to 5=not at all) (Patrick, 1999). I-QOL Total Score will be assessed using the responses from all 22 items. Three subscales will also be assessed:

- I-QOL Avoidance & Limiting Behaviors (Questions 1, 2, 3, 4, 10, 11, 13, and 20)
- I-QOL Psychosocial Impacts (Questions 5, 6, 7, 9, 15, 16, 17, 21, and 22)
- I-QOL Social Embarrassment (Questions 8, 12, 14, 18 and 19)

The I-QOL total and its subscale scores are computed by adding each non-missing item's response, subtracting the lowest possible score and dividing that sum by the possible raw score range. The scores are then transformed to have a range from 0 (poor quality of life) to 100 (maximum quality of life). The formula used to compute the transformed score is:

$$Scale\ Score = \frac{sum\ of\ the\ items - lowest\ possible\ score}{possible\ raw\ scale\ score\ range} * 100$$

The I-QOL Total Score will be set to missing if more than three items are left unanswered, and the subscales will be set to missing if more than one item is left unanswered.

3.6 PELVIC ORGAN PROLAPSE/URINARY INCONTINENCE SEXUAL QUESTIONNAIRE

The Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire (PISQ-IR) is a questionnaire designed to assess subscales separately for sexually active versus non-active subjects (Rockwood, 2013).

The following subscales are defined for sexually active subjects:

- Arousal and Orgasm (Questions 7, 8a, 10, and 11)
- Partner-Related (Questions 13, 14a, and 14b)
- Condition-Specific (Questions 8b, 8c, and 9)
- Global Quality Rating (Questions 19a, 19b, 19c, and 20a)
- Condition Impact (Questions 18, 20b, 20c, and 20d)

- Desire (Questions 15, 16, and 17)

The following subscales are defined for subjects that are not sexually active:

- Partner-Related (Questions 2a and 2b)
- Condition-Specific (Questions 2c, 2d, and 2e)
- Global Quality Rating (Questions 4a, 4b, 5a, and 6)
- Condition Impact (Questions 3, 5b, and 5c)

Responses to some questions must be reversed prior to scoring. The new values are obtained by subtracting the original response from the maximum value plus 1. For an item with 5 responses, this means a response of 1 becomes a 5, 2 becomes a 4, etc. The following questions must be reversed: 2 (a, b, c, d, e), 5 (a, b, c), 8 (b,c), 9, 11, 14 (a,b), 16, 17, 18, 19 (a,b,c). If Question 11 is missing and the checkbox is checked, a value of 5 will be used for the scoring.

The PISQ-IR subscale scores are computed by adding each non-missing item's response, subtracting the lowest possible score and dividing that sum by the possible raw score range. The scores are then transformed to have a range from 0 (worst) to 100 (best). The formula used to compute the transformed score is:

$$Scale\ Score = \frac{sum\ of\ the\ items - lowest\ possible\ score}{possible\ raw\ scale\ score\ range} * 100$$

These subscale scores require a minimum of two non-missing responses, except the Not Sexually Active - Partner-Related subscale which requires only one. If the requirement is not met, the subscale score will be set to missing.

3.7 DEVICE USE

Subjects randomized to the Vital Compact device are to use the device once per day for five days each week for the 12-week period. Subjects randomized to the itouch Sure Pelvic Floor Exerciser are to use the device once per day for the 12-week period. Subjects are to complete a daily record of device use, including their first in office session.

3.7.1 AVERAGE NUMBER OF SESSIONS PER WEEK

The average number of sessions per week is calculated as

$$Average\ Number\ of\ Sessions\ per\ Week = \frac{number\ of\ sessions\ used}{duration\ of\ treatment} * 7$$

where the number of sessions used is obtained from the daily record and the duration of treatment is the relative day of the Week 12 visit minus 1.

3.7.2 DEVICE USE – PERCENT OF ACTUAL USE

The percentage of actual use (through the usage period) is calculated as

$$Actual\ Use\ (\%)\ = \frac{number\ of\ sessions\ used}{number\ of\ sessions\ expected} * 100$$

where the number of sessions used is obtained from the daily record and the number of sessions expected is based on the relative day of the Week 12 visit. [For itouch Sure subjects, the number of sessions expected is the relative day of the visit minus 1; for Vital Compact subjects, it is the truncated value of 5/7*(relative day - 1)].

3.7.3 DEVICE USE – PERCENT OF TARGET USE

The percentage of target use (through Week 12) is calculated as

$$Target\ Use\ (\%)\ = \frac{number\ of\ sessions\ used}{number\ of\ sessions\ prescribed} * 100$$

where the number of sessions used is obtained from the daily record and the number of sessions prescribed is based on 12 weeks of usage. [For itouch Sure subjects, the number of sessions prescribed is 12*7=84; for Vital Compact subjects, it is 5*12=60).

4 CHANGES FROM STATISTICAL METHODS IN PROTOCOL

The protocol defined significant improvement as greater than 50% reduction in pad weight from baseline. This was clarified to be a 50% reduction in urine leakage, which is determined from the difference in pad weights pre and post testing.

The protocol planned to use both ITT and PP populations for the analyses of sensitivity, poolability, and subgroups. This was revised to only conduct these analyses on the ITT population.

The subgroup of mean intensity of the stimulation delivered during the 12-week treatment programme was only planned to be presented for urine leakage on the provocative pad weight test at Week 12. This was revised to also present this subgroup for the primary endpoint (significant improvement) at Week 12.

The subgroup of race was only planned to be presented for the primary endpoint at Week 12. This was revised to also present this subgroup for the change in urine leakage on the provocative pad weight test at Week 12.

The protocol provided ranges for the severity of urinary stress incontinence based on the 24-hour pad weight tests, but the ranges were incorrectly stated (1-hour pad weight test ranges were provided in error). This was revised to use the appropriate ranges per the O'Sullivan reference.

5 REFERENCES

Schafer, J.L. (1997) Analysis of Incomplete Multivariate Data. London: Chapman and Hall.

Yan X, Lee S, Li N Missing data handling methods in medical device clinical trials J. Biopharm. Stat. 2009: 19; 1085-1098

R.O'Sullivan, et al "Definition of Mild, Moderate and Severe Incontinence on the 24-hour pad test". BJOG, International Journal of Obstetrics and Gynecology. August 2004. Volume III, pgs. 859-862.

Patrick D, Martin M, Bushnell D, Yalcin I, Wagner T, Buesching D. Quality of life of women with urinary incontinence further development of the incontinence quality of life instrument (I-QOL). Urology 1999; 53(1): 71-6.

Rockwood TH, Constantine ML, Adegoke O, Rogers RG, McDermott E, Davila GW, Domoney C, Jha S, Kammerer-Doak D, Lukacz ES, Parekh M, Pauls R, Pitkin J, Reid F, Ridgeway B, Thakar R, Sand PK, Sutherland SE, Espuna-Pons M The PISQ-IR: considerations in scale scoring and development. Int Urogynecol J 2013. doi:10.1007/s00192-012-2037-z

Food Drug Administration Center for Devices and Radiological Health (2011) Guidance for Industry and FDA Staff - Clinical Investigations of Devices Indicated for the Treatment of Urinary Incontinence