STATISTICAL AND ANALYSIS PLAN

PROTOCOL NUMBER: D-FR-52120-236

PROTOCOL TITLE: A PHASE II, MULTICENTRE, DOUBLE-BLIND, RANDOMISED, PLACEBO-CONTROLLED, DOSE ESCALATION AND DOSE FINDING STUDY TO EVALUATE THE EFFICACY AND SAFETY OF DYSPORT IN VULVODYNIA PATIENTS

PROTOCOL VERSION AND DATE: 6.0 – 08 OCTOBER 2020

SAP Version	Date
Final Version 1.0	11 February 2021

STUDY NUMBER:	D-FR-52120-236
IND NUMBER	137038
PROTOCOL TITLE:	A PHASE II, MULTICENTRE, DOUBLE-BLIND, RANDOMISED, PLACEBO-CONTROLLED, DOSE ESCALATION AND DOSE FINDING STUDY TO EVALUATE THE EFFICACY AND SAFETY OF DYSPORT IN VULVODYNIA PATIENTS
SAP VERSION:	Final Version 1.0
SAP DATE:	11 February 2021

The undersigned agree that all required reviews of this document are complete, and approve this Statistical Analysis Plan:

Name	Company	Function	Date	Signature
PPD	IPSEN	PPD		e-signature
PPD	IPSEN	PPD		e-signature
PPD	Medpace	PPD		e-signature

History of Changes				
Old Version Number		Date Old Version	Date New Version	Reason for Change
Page	Section	Was	Is	

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

ADA:	Anti-drug Antibodies
AE:	Adverse Event
AESI	AEs of special interest
ALP:	Alkaline phosphatase
ALT:	Alanine aminotransferase
AST:	Aspartate aminotransferase
ATC:	Anatomic Therapeutic Class
ANOVA:	Analysis of Variance
ANCOVA:	Analysis of Covariance
CGI:	Clinical Global Impression
BMI:	Body Mass Index
BTX	Botulinum Toxin
BTX-A:	Botulinum toxin-A
CRF:	Case Report Form
CRO:	Clinical Research Organisation
DB:	Double-Blind
DLE:	Dose limiting event
DRC:	Data Review Committee
DMTS:	Dilator maximum tested size
e:	Electronic
EOS:	End of study
ICH:	International Council for Harmonisation
IMP:	Investigational Medicinal Product
MedDRA:	Medical Dictionary for Regulatory Activities
mFSFI:	modified Female Sexual Function Index
mITT:	modified Intent-To-Treat
mVPAQ:	modified Vulvar Pain Assessment Questionnaire
NRS:	Numeric Rating Scale
OL:	Open-Label
PCSA:	Potentially Clinically Significant Abnormalities
PGI:	Patient Global Impression
PGI-C:	Patient Global Impression of change in pain
PGI-S:	Patient Global Impression of severity of the pain
PHQ-9:	Patient Health Questionnaire-9

PP:	Per Protocol
PV:	Pharmacovigilance
PVD:	Provoked vestibulodynia
PVD1:	Primary PVD subjects
PVD2:	Secondary PVD subjects
SAP:	Statistical and Analysis Plan
SAE:	Serious Adverse Event
SAS®:	Statistical Analysis System®
SI:	Standard International
SDTM:	Study Data Tabulation Model
SOP:	Standard Operating Procedure
TEAE:	Treatment Emergent Adverse Event
WHO:	World Health Organization

1 INFORMATION TAKEN FROM THE PROTOCOL

1.1 Study objectives

1.1.1 Primary objective

Stage 1

The primary objective of the study is to assess the safety of increasing doses of Dysport in vulvodynia subjects with provoked vestibulodynia (PVD).

Stage 2

The primary objective of the study was to assess the efficacy of Dysport compared to placebo in vulvodynia subjects with PVD. However, the study was terminated before Stage 2 begins (please refer to Section 1.3.5).

1.1.2 Secondary objectives

The secondary objectives of the study are as follows:

Stage 1

- To assess the efficacy of increasing doses of Dysport compared to placebo in vulvodynia subjects with PVD
- To determine the doses of Dysport to be administered in Stage 2
- To assess the sensitivity of the endpoints used in Stage 1

Stage 2

- To define the optimal doses of Dysport with an acceptable benefit/risk profile for the treatment of vulvodynia with PVD
- To assess effect of Dysport on:
  - Vulvar pain
  - Use of pain rescue medication (type, dose and frequency)
- To determine if primary and secondary PVD subjects benefit from Dysport administration
- To assess the time to retreatment
- To assess the safety of Dysport

1.1.3 Exploratory objectives

The exploratory objectives of the study are as follows:

- To assess effect of Dysport on:
  - Clinical Global Impression (CGI) of the treatment effect (assessed by the investigator)
  - Patient Global Impression of severity of the pain (PGI-S) and Patient Global Impression of change in pain (PGI-C)
- Emotional response, cognitive response and associated life interference as assessed on modified Vulvar Pain Assessment Questionnaire (mVPAQ) subscales
- Sexual function
- To assess the efficacy of Dysport on the pelvic floor muscles pressure
- To assess the efficacy of Dysport on depression
- To assess the efficacy of Dysport on the quality of life

1.2 Study design

This is a Phase II multicentre, double-blind (DB), randomized, placebo-controlled, dose finding study to define the optimal doses of Dysport and evaluate its efficacy and safety compared to placebo in vulvodynia patients with PVD. Study treatment was injected at 5 needle insertion points at 10 injection sites in the vestibular area. The study consisted of a dose escalation stage (Stage 1) and a dose expansion stage (Stage 2). The purpose of the dose escalation was to determine the Dysport doses to be further investigated in Stage 2. Both Stage 1 and Stage 2 consisted of a DB period (treatment cycle 1; Dysport or placebo) followed by an open-label (OL) treatment period (treatment cycles 2 to 4; all subjects receive Dysport).
However, the study was terminated before Stage 2 begins (please refer to Section 1.3.5).

1.3 Methods and procedures

1.3.1 Schedule of assessments

The schedule of procedures and assessments during the study is presented in section 5.1 of the protocol.

1.3.2 Planned sample size

In Stage 1, the expected sample size was 10 evaluable subjects for PVD2 (secondary PVD subjects, i.e. having a past history of pain free intercourse or insertion of any object >1 cm diameter) cohort to be randomized and treated in a 4:1 allocation ratio (Dysport or Placebo) and 8 evaluable subjects for PVD1 (primary PVD subjects, i.e. having life-long PVD) cohort 4 treated in a 3:1 allocation ratio (Dysport or Placebo). After Cohort 3 (PVD2) and Cohort 4 (PVD1), the expected sample size was 10 evaluable subjects for additional PVD1 and PVD2 cohorts and treated in a 4:1 allocation ratio (Dysport or Placebo).
Subjects who discontinue prior to Week 6 for reasons unrelated to safety may be replaced at the discretion of the investigator and sponsor. As this is an exploratory stage, no formal power was performed. However, with 8 active subjects per PVD2 dose cohort and if there are 0 dose limiting events (DLEs) in a cohort, there is at least 80% probability that the true incidence of DLEs at the corresponding dose will not be greater than 18%. With 16 active subjects in PVD1 and PVD2 dose cohorts and if there are 0 DLEs in a cohort, there is at least 80% probability that the true incidence of DLEs at the corresponding dose will not be greater than 9%.

1.3.3 Method of Randomisation

In stage 1, for the PVD2 cohorts, subjects were randomised with a 4:1 (Dysport: placebo) ratio. For the first PVD1 cohort, subjects were randomised with a 3:1 (Dysport: placebo) ratio for Dysport 400 U and then for further PVD1 cohorts, subjects were randomised at 4:1 for Dysport 500 U and up to a maximum of 800 U.
In an event, when the last subject in a cohort was being randomised, if an additional subject had been screened, this subject may be accepted in the cohort and be randomised.

1.3.4 Validation scale study D-FR-52120-252

The psychometric properties of the dilator-induced pain Numeric Rating Scale (NRS), dilator size 6 insertion pain NRS, pain on intercourse NRS, mVPAQ, and modified Female Sexual Function Index (mFSFI) were to be evaluated using data collected from all subjects randomized in Stage 1 and Stage 2. The planned analyses of these data are described in the Psychometric Analysis Plan of validation study D-FR-52120-252. However, these analyses were cancelled following the decision to prematurely terminate the study.

1.3.5 Data review Committee and planned Analyses

Data Review Committee

A DRC has been established to make recommendations for dose escalation and dose de-escalation or study termination decisions in Stage 1. It was planned that (a) if no subject reports a clinically significant AE (defined in the protocol section 3.1.1.1) in a cohort, the DRC will meet once all PVD2 and PVD1 evaluable subjects of a dose level have reached Week 6 to review the unblinded efficacy and safety data and decide the dose for the next cohort or proceed to Stage 2; (b) If any subject reports a clinically significant AE, then further enrolment into the cohort as well as retreatment in open label cycles at the current dose will be suspended and an ad hoc DRC meeting will be scheduled to review the data and make a recommendation with regards to further conduct of the cohort and of Stage 1.

Details related to DRC membership, responsibilities, procedures, role and remit are described in the DRC charter.

Early Study Termination Decision

Following the DRC meeting organized to review safety and efficacy data of cohorts 1 to 6 on the 16 OCT 2020, the DRC sent the following recommendation:

"No safety concerns noted in review of all available data. Efficacy cannot be assessed due to the small sample size of cohorts. The DRC would recommend that Stage 1 is not used to make an efficacy assessment, therefore there is no justification to increase the dose, nor to move to Stage 2."

Ipsen subsequently decided to close the study early and to run the final analysis based on available Stage 1 data. Subjects in the study are followed until 12 weeks after the last dose as described in the protocol section 3.9.

Final Analysis

The final analysis will be carried out once all subjects from cohorts 1 to 6 have completed end of study evaluation, after the database lock. All unblinded safety and efficacy data available at the time of the database lock will be analysed.

No control of the overall type I error rate is planned as only exploratory descriptive statistics are planned.

2 SUBJECT POPULATIONS (ANALYSIS SETS)

Screened population

The screened population is all subjects screened (i.e. who signed the informed consent).

Randomized population
Randomized population is all subjects randomized (i.e. who were randomly allocated to a treatment group).

modified Intent-To-Treat population
The modified Intent-to-Treat (mITT) population is all randomized subjects who received at least one IMP administration and had data for the vaginal dilator induced pain as reported on an 11-point NRS [using the Dilator Maximum Tested Size (DMTS) reported at Baseline] at Baseline and Cycle 1-Week 6 visit (i.e. at least one pain score for one dilator size, which enables an imputed pain score of 10 at DMTS if needed).
The mITT population will be analyzed using subjects as randomized.

Active Treatment Population
The Active Treatment Population is all randomized subjects who received at least one dose of Dysport (including only partial administration) during the DB and/or OL period.

Safety
The safety population is all randomized subjects who received at least one dose of IMP administration (including only partial administration).
The safety population will be analyzed using subjects as treated (i.e. according to actual treatment received).

3 STATISTICAL METHODS

3.1 Statistical analysis strategy

The statistical analyses will be performed in accordance with the ICH E9 guideline and will be based on the pooled data from the individual study sites, unless otherwise stated.
Medpace will perform the statistical analysis of the efficacy, safety, and antibodies data.
Overall, the analysis strategy is to evaluate safety and efficacy of Dysport compared to Placebo after single administration (DB period), and to evaluate efficacy and safety over repeated Dysport treatments, using analyses by Active Treatment Cycle.
Listings will be presented by subject for DB and OL periods together while the tables associated with the DB period will be separated from the tables assessing repeat treatment with Dysport.

Only descriptive statistics of Stage 1 data will be provided.

Double-Blind Period (Stage 1 Cycle 1) Analyses
For the analysis of the DB period, tables will be displayed using the following treatment group (DB treatment group) labels, in the following order:

- Placebo: placebo groups from all cohorts will be pooled, including for mVPAQ pain, mVPAQ life interference and mFSFI pain domain of cohort 1 despite the change in recall period and time of assessments (see protocol version 3.0).
- Dysport
    - Groups by dose (Dysport 100U, 300U, 400U, 500U, etc.)
    - All Dysport

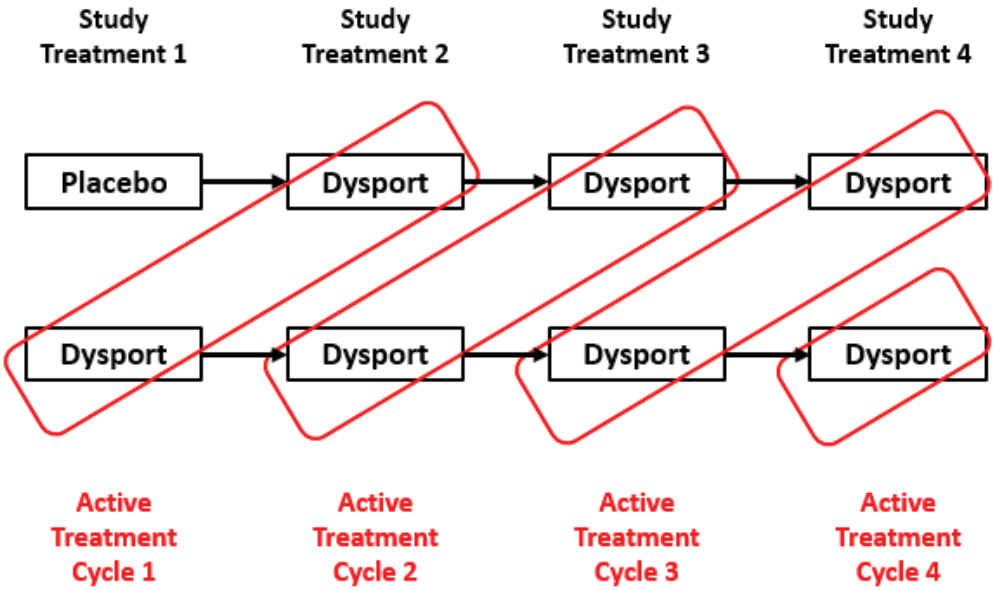
Note: Cohort 1 (100U), 2 (300U), 3 (400U) and 5 (500U) included PVD2 subjects only. Cohort 4 (400U) and 6 (500U) included PVD1 subjects only.

Repeat treatment: Active Treatment Analyses

To assess efficacy and safety data over repeat treatments, data will be analysed using the Active Treatment Cycle approach as described in Figure 1 below. These analyses will be performed on the Active Treatment population:

- for subjects treated with Dysport in the DB period, the first cycle in the DB period will be the first Active Treatment Cycle.
- for subjects treated with placebo in the DB period, the first cycle in the OL period will be the first Active Treatment Cycle.

Figure 1 Active Treatment Cycle Analysis



Three types of summary tables will be produced:

- Summary by number of active treatment cycles: these tables will present the long-term efficacy and safety data with all doses combined, and will display subjects with at least 1 (2, 3 and 4) active treatment cycles.
- Summary by dose received at each active treatment cycle: to evaluate the safety of each dose at a given cycle.
- Summary over the whole trial (restricted to Week 6 and 12 from DB and from first OL period) by the randomised treatment group (DB treatment group) may be provided for some efficacy analyses.

The type of summary table(s) used for each endpoint are described in sections 3.8 and 3.9.

Descriptive Statistics

All raw and derived variables will be listed and described using summary statistics. For categorical variables, summary statistics will be displayed by frequency counts and percentages by category. The missing category will be presented if there is at least one missing data in a treatment group. Except otherwise specified, subjects with missing data will be included in the calculation of percentages. For quantitative

variables, summary statistics will be displayed using descriptive statistics by number of observations, mean, standard deviation (SD), median, minimum and maximum.

3.1.1 Multiplicity

No multiple testing will be performed.

3.1.2 Significance testing and estimation

As this is a descriptive safety study, no formal statistical testing will be carried out based on Stage 1 data.

3.1.3 Withdrawal visit and end of study visit

Subjects who have withdrawn early from the study have their last assessment (during early withdrawn visit) mapped as visit 90 in the Study Data Tabulation Model (SDTM). By convention, for these subjects, the visit number will be reassigned to the next empty visit number for analysis (i.e. if a subject has only Visit 1, 2, 3 and 90 entered in the database, the visit number 90 will be reassigned to visit number 4). End of study visits will be described separately as "End of Study (Cycle X)".

3.2 Subject disposition and population

The following disposition summaries and listings will be provided:

- Number and percentages of randomized subjects per country and site by DB treatment group.
- Number and percentages of subjects screened, screen failed and reason for screen failure, randomized, treated, completed, withdrawn and reason for withdrawal by DB treatment group.
- Listing of date of visit.
- Listing of screen failure subjects.
- Listing of withdrawn subjects (on the randomised population).

The following population summaries and listings will be provided:

- Number and percentage of subjects in each analysis population with reasons for exclusion from each population, for the DB treatment group and overall, based on all randomized subjects.
- Listing of randomization details.
- Listing of subjects who don't meet eligibility criteria (violated inclusion criteria, fulfilled exclusion criteria).
- Listing of subjects including flag for each analysis population and reason for exclusion from each population.

3.3 Protocol Deviations

A list of major protocol deviations and any action taken have been defined in the Protocol Deviation Plan. Subjects with major protocol deviations will be determined before database lock and unblinding of the study. A final review will be done during the blind data review and documented in the protocol deviation listing. Confirmation of major deviations related to treatment and dose will be done after unblinding.
Number and percentage of subjects with major protocol deviations will be summarized by deviation category for all randomized subjects. A Listing of major

protocol deviations will also be provided. COVID-19 related protocol deviations will be flagged.

3.4 Demographic and Baseline characteristics

All demographic and Baseline characteristics (including the Pain Catastrophizing Scale total score) will be listed by treatment group, and subject for the randomised population. The Pain Catastrophizing Scale total score is calculated as the sum of the 13 items, ranging from 0 to 52.
Summary statistics will be provided for demographic and Baseline characteristics, by DB treatment group, for the Randomized, and mITT populations.
Q-tip test results will be listed for all screened subjects.
No statistical comparison of the treatment groups will be provided.

3.5 Medical and surgical history

Medical and surgical history, non-drug therapies and surgical procedures will be coded using the latest version of MedDRA at the time of the database lock.
Listings will present the primary system organ class, preferred term and verbatim text. The listings will be sorted by DB treatment group, subject, primary system organ class, preferred term and verbatim text.
A frequency table of the number and percentage of subjects will be provided for all medical and surgical history by primary system organ class and preferred term. The table will be presented by DB treatment group for the randomised population.

3.6 Prior and concomitant therapies

Prior and concomitant therapies will be coded using WHO-Drug Dictionary, the latest version at the time of database lock. The therapeutic class will correspond to the second level of ATC code.

Medication, non-drug therapies, physiotherapies, sex therapies and surgical procedures start and stop dates will be compared to the date of the first IMP administration to allow classification as either Prior (P) only or Concomitant (PC, C):

Prior (P)	Start and stop dates prior to the date of the first IMP administration.
Prior and Concomitant (PC)	Start date before the date of the first IMP administration and stop date on or after the date of the first IMP administration.
Concomitant (C)	Start date on or after the date of first IMP administration.

Summary tables on prior medications/non-drug therapies/surgical procedures will include “P” only, summary tables on concomitant medications/non-drug therapies/surgical procedures will include “C” and “PC”.

A treatment will be considered concomitant to a cycle if the start date was on or before the IMP administration for the cycle and the stop date was on or after the IMP administration for the cycle. A treatment can be concomitant under more than one treatment cycle, but it will be considered as concomitant only once in the overall period.

The following summaries will be provided:

- number and percentage of subjects with prior medication (category P) by ATC class and Preferred Name (PN, ATC level 2),
- number and percentage of subjects with concomitant medication (categories PC, C) by ATC class and PN (ATC level 2),
- number and percentage of subjects with prior non-drug therapy (category P) by SOC and PT,
- number and percentage of subjects with concomitant non-drug therapy (categories PC, C) by SOC and PT,
- number and percentage of subjects with prior pelvic floor physical therapy (category P) by SOC and PT,
- number and percentage of subjects with concomitant pelvic floor physical therapy (PC, C) by SOC and PT,
- number and percentage of subjects with prior sex therapy (category P) by SOC and PT,
- number and percentage of subjects with concomitant sex therapy (categories PC, C) by SOC and PT,
- number and percentage of subjects with concomitant surgical procedures by primary SOC and PT.

All these tables will be presented during the DB period by DB treatment group for the randomized population. In addition, tables related to concomitant medications/ therapies/ surgical procedures will be presented during Active Treatment period all doses combined by Active Treatment Cycle and overall, for the Active Treatment population.

Listings will be provided for all the summaries listed above with the therapeutic class, preferred term/name and verbatim text. The level of concomitance will also be presented (P, PC, C).

A listing of prohibited concomitant medications and a listing of pain rescue medications will also be provided.

The listings will be sorted by DB treatment group, subject, chronological start date, primary system organ class/therapeutic class, preferred term/name and verbatim name.

3.7 Subject compliance

A listing will be presented for drug administration (volume administered, specified difficulties during drug administration, injection sites, date) at all treatment cycles by DB treatment group and subject.

3.8 Efficacy analysis

The efficacy endpoints will be summarized using the analysis strategy in Section 3.1. Descriptive statistics will be provided for:

- the DB period, by DB treatment group for the mITT population.
- the Active Treatment Cycle, on the Active Treatment population: by number of active treatment cycles (all doses combined and/or by dose). The mITT population will be the primary population for analysis of efficacy data. Week 6 and Week 12 will be the timepoints of primary interest.

3.8.1 Primary efficacy analysis

Not applicable as the primary objective of Stage 1 is related to safety.

3.8.2 Secondary efficacy analyses

Vestibular pain using a vaginal dilator

A set of 8 vaginal dilators of increasing diameter (#1 being the smallest and #8 being the largest) will be used to provoke pain to allow assessment of the vestibular pain intensity in each subject. Subjects will be required to rate the pain for each dilator size on an 11-point NRS at each visit. Pain score will be entered by the investigator (or delegated site staff) in the eCRF.

Based on the subjective pain threshold, the largest sized dilator that the subject accepts to test, will be defined as the Dilator Maximum Tested Size (DMTS) that will be established at the Baseline Visit. The dilator test will be used to assess the change from Baseline in the vaginal dilator induced pain as reported on an NRS (using the DMTS reported at baseline) as well as the dilator size that provokes maximum pain.

If a dilator size is not tested due to pain intensity, the pain will be scored at 10. If the test is not done due to other reasons, then the pain score will be considered as missing.

Endpoints and corresponding analyses are described in Table 1.

Table 1 Dilator Induced Pain - Endpoints and Corresponding Analyses

Endpoint	Analyses
Mean change from Baseline to each post-treatment visit in the vaginal dilator induced pain as reported on the 11-point NRS[1]	Convention: According to the mITT Population definition, there will be no missing data for Week 6. For Week 12, the missing values will be kept as missing. - descriptive statistics: actual values and change from baseline values will be presented at all timepoints (over the double-blind and open-label periods). A plot of mean changes over time from baseline in pain scores will be provided by treatment group for DB period. Active Treatment Cycle: - A summary by number of active treatment cycles. - A plot of mean changes over time from baseline in pain scores will be provided by randomized treatment group for Week 6 and Week 12 over both of the DB and first OL period.
Proportion of subjects who reported at least a 50% decrease in vaginal dilator induced pain as reported on the 11-point NRS[1] at each post-treatment	Convention: According to the mITT Population definition, there will be no missing data for Week 6. For Week 12, the missing values will be kept as missing. - descriptive statistics will be presented at all timepoints.

visit (repeat for 30% decrease) Proportion of subjects who reported at least 2-point decrease from baseline in vaginal dilator pain as reported on the 11-point NRS[1] at each post-treatment visit	Active Treatment Cycle: - A summary by number of active treatment cycles will be used.
Mean change from Baseline to each post-treatment visit in the dilator size that provokes maximum tolerated pain	- descriptive statistics: actual values and change from baseline values will be presented at all timepoints. - Number and percentage of each size will be provided at all timepoints. Active Treatment Cycle: - A summary by number of active treatment cycles will be used.
Mean change from Baseline to each post-treatment visit in the composite score for the vaginal dilator induced pain and dilator size	Composite score derivation: The composite score for the vaginal dilator induced pain and dilator size is the sum of all pain measurements across the full range of dilator sizes. In the construction of this endpoint, the pain score will be 10 for any dilator size that is beyond the DMTS. As described in section 3.10.1, there should be at least 6 pain scores (recorded by the investigator) to calculate the composite score. When there is a dilator size score missing, pain score will be imputed by the average of the previous (smaller dilator size) and the following (larger dilator size) valid values. For example, if the pain score for dilator size 4 is missing, then it will be imputed by the average of the pains scores for dilator size 3 and size 5 when both of them are available. - descriptive statistics: actual values and change from baseline values will be presented at all timepoints. Active Treatment Cycle: - A summary by number of active treatment cycles will be used.

[1] Using DTMS at baseline.

Pain during insertion of vaginal dilator number 6 size

Cohort 1: once a week during the Screening period (Day -14 to Day 1), and once a week during the month (28 days) prior to the next planned visit, all subjects are to rate pain following insertion of the number 6 vaginal dilator (11-point NRS) and to record details of associated pain medication if consumed. Data are recorded in the patient's eDiary.

Cohort 2 onwards: once a week during the 2 weeks preceding the next planned visit, all subjects are to rate pain following insertion of the number 6 vaginal dilator (11-point NRS) and to record details of associated pain medication if consumed. Data are recorded in the patient's eDiary.
Endpoints and corresponding analyses are described in Table 2.

Table 2 Dilator #6 - Endpoints and Corresponding Analyses

Endpoint	Analyses
Mean change from Baseline to each post-treatment visit in pain during insertion of vaginal dilator number 6 size as reported on the 11-point NRS	Convention: Average of the values over the last 2 weeks preceding the visit. The average will be missing if and only if the two pain scores are both missing. This rule will be applicable for all cohorts including cohort 1. - descriptive statistics: actual values and change from baseline values will be presented at all timepoints. Active Treatment Cycle: - A summary by number of active treatment cycles will be used.
Proportion of subjects who reported at least a 50% decrease in pain during insertion of vaginal dilator number 6 size as reported on the 11-point NRS at each post-treatment visit (repeat for 30% decrease)	Convention: Average of the values over the last 2 weeks preceding the visit. The average will be missing if and only if the two pain scores are both missing. This rule will be applicable for all cohorts including cohort 1. - descriptive statistics will be presented at all timepoints. Active Treatment Cycle: - A summary by number of active treatment cycles will be used.
Proportion of subjects who reported at least 2-point decrease from baseline at each post-treatment visit	
Proportion of subjects taking pain rescue medication associated with insertion of	- descriptive statistics: number and percentage of subjects using pain rescue medication within the 2 weeks period preceding each visit, in subjects with non-missing pain score value in the corresponding period.

vaginal dilator number 6 size.	Active Treatment Cycle: - A summary by number of active treatment cycles will be used. Subjects having taken pain rescue medication associated with insertion of vaginal dilator will be listed, along with period, medication (primary system organ class, preferred term and verbatim text), name, number of pills, dosage of each pill, route, date, taken before/after the dilator insertion and pain during the dilator insertion in the corresponding week.

Frequency of intercourse and pain during intercourse

Cohort 1: Each day during the Screening period (Day -14 to Day 1), and during the month (28 days) preceding the next planned visit, subjects are to record in the eDiary if they had intercourse, the number of intercourse instances, corresponding pain during each intercourse instance (using an 11-point NRS) and details of pain rescue medication, if taken, to prevent or treat vestibular pain.

Cohort 2 onwards: Each day during the Screening period and during the 2-week period preceding the next planned visit, subjects are to record if they had intercourse, the number of intercourse instances, corresponding pain during each intercourse instance on an 11-point NRS and details of pain rescue medication, if taken, to prevent or treat vestibular pain.

Intercourse diary entry with no intercourse in the previous 24 hours and pain score entered as 0 will be considered as no intercourse in the previous 24 hours. Entry with no intercourse in the previous 24 hours and pain score different from 0 will be considered as an intercourse entry with the entered pain score.

Endpoints and corresponding analyses are described in Table 3.

Table 3 Intercourse - Endpoints and Corresponding Analyses

Endpoint	Analyses
Mean change from baseline to each post-treatment visit in the pain during intercourse as reported on the 11-point NRS	Conventions: Analysis of pain will be done on the subjects who had intercourse in the corresponding period. Average pain score over the week preceding the visit: The weekly average will be calculated as the average of the score over the last week (7 days) prior to the visit, for all cohorts (including cohort 1). The weekly average will be calculated if at least 1 pain score has been recorded over the considered week. Average pain score over the 2 weeks preceding the visit: The 2-week average will be calculated as the average of the score over the last 2 weeks (14 days) prior to the visit, for all cohorts (including cohort 1). The 2-week average will be calculated if at least 1 pain score has been recorded in the considered 2-week period.

	- descriptive statistics: actual values and change from baseline values will be presented at all timepoints. In the same table, number of subjects with or without partner at each timepoint will be also presented. Active Treatment Cycle: - A summary by number of active treatment cycles will be used.
Mean change from baseline to each post-treatment visit in the number of intercourse instances in subjects with partners	Conventions: Analysis will be done on the subjects who have a partner at baseline and at the specific visit (DB Week 6, or Week 12). Number of intercourse instances over the week preceding the visit The number of intercourse instances over the week preceding the visit will be calculated as the total number of intercourse instances over 7 days prior to the visit. There should be at least 1 eDiary input recorded in the week for calculating the number of intercourse instances over the period, otherwise it will be considered missing. Number of intercourse instances over the 2 weeks preceding the visit The number of intercourse instances over the 2 weeks preceding the visit will be calculated as the total recorded number of intercourse instances over 14 days prior to the visit. There should be at least 1 eDiary input recorded in the period for calculating the number of intercourse instances over the period, otherwise it will be considered missing. - descriptive statistics: actual values and change from baseline values will be presented at all timepoints. Active treatment cycle: - A summary by number of active treatment cycles will be used.
Proportion of subjects having taken pain rescue medication associated with intercourse	- descriptive statistics: number and percentage of subjects using pain rescue medication within the 1-week/2-week period preceding each visit, in subjects with non-missing pain score value in the corresponding period. Active Treatment Cycle: - A summary by number of active treatment cycles will be used.

	Subjects having taken pain rescue medication associated with intercourse will be listed, along with period, medication (primary system organ class, preferred term and verbatim text), name, number of pills, dosage of each pill, route, date, taken before/after the intercourse instance and pain during the intercourse in the corresponding week.

3.8.3 Exploratory efficacy analyses

Clinical Global Impression (CGI) and Patient Global Impression (PGI)

Endpoints and corresponding analyses are described in **Error! Reference source not found.**.
CGI was assessed by the investigator using a 7-point Likert scale during the visit.
PGI-S and PGI-C: Subjects were to assess PGI-S and PGI-C in the eDiary one week prior to the next planned visit and at the visit. The outcome for the visit will be calculated as the average of a) the average of the entries on the date of the visit (in case of several entries on the day of the visit) and b) the average of the entries within the week prior to the visit (in case of several entries).

Table 4 CGI and PGI - Endpoints and Corresponding Analyses

Endpoint	Analyses
Mean score of CGI	- descriptive statistics will be presented at all timepoints. Active treatment cycle: - A summary by number of active treatment cycles will be used.
Mean change from baseline to each post-treatment visit in the PGI-S	- descriptive statistics: actual values and change from baseline values will be presented at all timepoints. Active treatment cycle: - A summary by number of active treatment cycles will be used.
Mean PGI-C at each post-treatment visit	- descriptive statistics will be presented at all timepoints. Active treatment cycle: - A summary by number of active treatment cycles will be used.

mVPAQ and mFSFI

Endpoints and corresponding analyses are described in Table 5.
mVPAQ questionnaire is provided in protocol Appendix 1. It contains 6 subscales: pain severity, emotional response, cognitive response, life interference, sexual

function interference, self-stimulation/penetration interference. The subscale score is the average of the item scores under the subscale.
For mVPAQ, the analysis of life interference subscale will be done with and without the item 'ability to fall asleep'. The analysis of pain severity subscale will be done with and without the items 'unpleasantness' and 'distress'. All these items will be included/excluded together when calculating subscale score.
For mVPAQ pain subscale, life interference subscale and mFSFI pain domain: Subjects were to record in the e-diary the data related to the completion of these questionnaires in the eDiary one week prior to the next planned visit and at the visit. The subscale scores for the visit will be calculated as the average of a) the average of the subscale scores entered on the date of the visit and b) the average of the subscale scores entered within the week prior to the visit.For other mVPAQ subscales and mFSFI domains, data were recorded by the subject at the site during the visit.

Table 5 mVPAQ and mFSFI - Endpoints and Corresponding Analyses

Scales	Endpoints	Analyses/ Conventions
Modified Vulvar Pain Assessment Questionnaire (mVPAQ) – Pain severity subscale	Change from baseline in each question and subscale score	The score for each question ranges from 0 to 4 and a mean score is computed. - descriptive statistics: actual values and change from baseline values will be presented at all timepoints.
Modified Vulvar Pain Assessment Questionnaire (mVPAQ) – Life interference subscale	Change from baseline in each question and subscale score	The score for each question ranges from 0 to 4 and a mean score is computed. - descriptive statistics: actual values and change from baseline values will be presented at all timepoints.
Modified Vulvar Pain Assessment Questionnaire (mVPAQ) – other subscales	Change from baseline in each subscale score	The score for each question ranges from 0 to 4 and a mean score is computed. - descriptive statistics: actual values and change from baseline values will be presented at all timepoints.
modified Female Sexual Function Index (mFSFI) – pain domain	Change from baseline in each question and total score	See Table 6 for domain and total score calculation. - descriptive statistics: actual values and change from baseline values will be presented at all timepoints.
Modified Female Sexual Function Index (mFSFI) – other domains	Change from baseline total score in each domain	See Table 6 for domain and total score calculation. - descriptive statistics: actual values and change from baseline values will be presented at all timepoints.

modified Female Sexual Function Index (mFSFI) – all domains	Change from baseline in mFSFI total score.	See Table 6 for domain and total score calculation. - descriptive statistics: actual values and change from baseline values will be presented at all timepoints.
Pelvic floor muscles pressure – Resting and Squeezing	Mean change from baseline in the pelvic floor muscles pressure for the 2 conditions: - Resting vaginal pressure - Maximal 'squeeze' pressure, and the difference between the 2 conditions (squeeze minus rest condition)	Conventions: The mean of the 3 measurements will be used. The mean will be missing if and only if the 3 measurements are missing. An assessment will be invalidated if the measurement is not accurate as specified by the investigator or if subject had a vaginal delivery. Difference of the mean between Squeeze and rest conditions will be calculated. - descriptive statistics: actual values and change from baseline values will be presented at all timepoints.
Patient Health Questionnaire (PHQ-9)	Change from baseline in the PHQ-9 total score	See Appendix 3 from protocol for total score definition. PHQ-9 questionnaire comprises 9 items, each of which is scored 0 to 3, providing a 0 to 27 total score. (If several entries are available on the date of visit, the average of the total score will be used) - descriptive statistics: actual values and change from baseline values will be presented at all timepoints.
36-Item Short Form Health Survey (SF-36)	Change from baseline in all 8 domains scores, and the 2 components (physical and mental)	The eight dimensions of SF-36 and the Physical and Mental component will be derived by OPTUM. - descriptive statistics: actual values and change from baseline values will be presented at all timepoints.

Table 6. mFSFI Score Calculation

Domain	Questions	Score Range	Factor	Minimum Score	Maximum Score
Desire	1, 2	1 – 5	0.6	1.2	6.0
Arousal	3, 4, 5, 6	0 – 5	0.3	0	6.0
Lubrication	7, 8, 9, 10	0 – 5	0.3	0	6.0
Orgasm	11, 12, 13	0 – 5	0.4	0	6.0
Satisfaction	14, 15, 16	0 (or 1) – 5	0.4	0.8	6.0
Pain	17, 18, 19	0 – 5	0.4	0	6.0
			Full Scale Score Range	2.0	36.0

Note: The individual domain scores and full scale (overall) score of the FSFI can be derived from the computational formula outlined in the table above (Rosen etc 2000). For individual domain scores, add the scores of the individual items that comprise the domain and multiply the sum by the domain factor. Add the six domain scores to obtain the full-scale score. It should be noted that within the individual domains, a domain score of zero indicates that the subject reported having no sexual activity during the recall period.

3.9 Safety

For the DB period, summary tables by DB treatment group will be provided on the Safety population.
For the Active Treatment Cycle, summary tables by number of active treatment cycles will be provided on the Active Treatment population. Additionally, a summary table will be also provided by dose received at each active treatment cycle on the Active Treatment population.

The summary tables will follow the strategy specified in Section 3.1.

3.9.1 Extent of exposure

To describe the duration of exposure, three durations of exposure will be calculated.

- Duration of study treatment exposure will be defined in weeks as:
  (The date of last visit attended – treatment cycle 1 injection date + 1) / 7.
- Duration of treatment cycle will be defined in days as:
  For subjects who are retreated at the end of the cycle, the duration of treatment cycle is calculated as (date of retreatment) - (treatment injection date).
  For subjects who are not retreated at the end of the cycle, the duration of treatment cycle is calculated as (last attended visit date of the treatment cycle) – (treatment injection date) + 1.
- Duration of Active Treatment exposure will be defined in weeks as:
  (The date of last visit attended – first Dysport injection date + 1) / 7.

The following summary tables will be provided:

- Description of study treatment exposure by DB treatment group + overall subjects on the safety population for DB period and overall subjects for the whole active treatment periods.
  Study treatment exposure will be described both as a quantitative and qualitative variable.

- Description of duration of treatment cycle, for each treatment cycle, dose received of each treatment cycle and overall subjects from the safety population.
- Description of duration of treatment cycle, by dose received at the cycle and all doses combined on the active treatment population.
  Active Treatment exposure will be described both as a quantitative and qualitative variable for each period of exposure.
- A summary table with the total number of subjects who received 1, 2 or 3 active treatment cycles and the total number of subjects who received at least 1, 2 or 3 active treatment cycles, all doses combined on Active treatment population and by DB treatment group + overall subjects.

To describe the doses received, the following summary tables will be provided:

- Description of the actual dose received in each cycle by DB treatment group and overall subjects from the safety population.
  The number and percentage of subjects receiving the actual dose will be provided.
- Description of the actual dose at each active treatment cycle, by treatment received for the active treatment cycle and overall subjects from the safety population.
  The number and percentage of subjects receiving the actual dose will be provided for each active cycle.
- A listing of exposure data will also be provided.

3.9.2 Adverse events

Adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) latest version (23.0 or later) and presented by MedDRA preferred term and system organ class.
AEs will be classified as treatment-emergent AEs (TEAEs) according to the rules below:

- Events with start date (and time) on or after the date (and time) of first IMP administration.
- Events whose severity worsens on or after the date (and time) of first IMP administration.

A TEAE that occurs under a particular treatment cycle and worsens in intensity under a different treatment cycle will be considered as treatment emergent under the two treatment cycles. In the event of another occurrence of a TEAE with a change in intensity under a different treatment cycle, the subject will also be counted in that different treatment cycle. A given AE will be assigned to the dose received prior to the onset of the AE in the designated treatment cycle.
An overall summary table of all adverse events will be presented (as described in Section 3.9) and include the following:

- Subjects with any TEAEs,
- Subjects with any TEAEs by maximum intensity,
- Subjects with any TEAEs related to IMP,
- Subjects with any TEAEs related to IMP by maximum intensity,
- Subjects with any TEAEs leading to drug withdrawal,
- Subjects with any SAEs,

- Subjects with any TEAEs leading to death,
- Subjects with any SAEs related to IMP,
- Subjects with any TEAEs of special interest (AESIs).

This summary table will be also be provided by dose received at each active treatment cycle.

All TEAEs will be summarized with the number and percentage of subjects with adverse events classified by primary system organ class and preferred term. The number of occurrences of a TEAE will also be presented. This summary table will also be provided by dose received at each active treatment cycle.

Summary tables will also be presented by SOC and PT for SAEs, TEAEs by intensity, TEAEs by causality, and non-serious TEAEs with at least 2 subjects. TEAEs will also be summarized by decreasing overall PT frequency

If the intensity or causality is missing for a TEAE, it will be considered as missing in the summary tables. In the event of multiple TEAE being reported by the same subject within the same SOC/PT, the maximum intensity (severe > moderate > mild > missing) and the most serious causality (related > not related > missing) will be chosen.

Listings will be presented and sorted by DB treatment group, subject ID, start date of AE, primary system organ class, preferred term and verbatim text for all adverse events recorded during the study. Other information collected, such as end date and outcome of AE, will be provided in the listings.

Listings of SAE, adverse events leading to discontinuation from the study, clinically significant events, adverse events leading to drug withdrawal and death will also be presented.

TEAEs will be identified in the adverse events listings. Pre-treatment AEs and clinically significant AEs will also be flagged in the AE listings.

Adverse Events of Special Interest (AESIs)

Adverse events of special interest (AESIs) for Dysport in this study are defined as:

- Any TEAE that suggests a possible remote spread of effect of the toxin, or
- Any TEAE related to urinary incontinence or faecal incontinence, or
- Any TEAE assessed as a potential hypersensitivity reaction,

where there is a reasonable possibility of causal relationship with treatment with Dysport, regardless of the investigator reported causality.

The method described below is used for identification and analysis of events that may be qualified as TEAEs that suggest a possible remote spread of effect of the toxin, events related to incontinence or faecal incontinence or events suggestive of hypersensitivity like reactions.

TEAEs due to possible remote spread of the effects of Dysport will be identified using the list of MedDRA preferred terms (PTs) compatible with the mechanism of action of BTX-A and based on the recommendations from the Committee for Medicinal Products for Human Use (CHMP) and the Food and Drug Administration (FDA). TEAEs potentially representing hypersensitivity reactions will be identified using the Standardised MedDRA Query (SMQ) (narrow search query) for hypersensitivity reactions. TEAEs that are possibly related to urinary incontinence or faecal incontinence will also be identified using the list of MedDRA preferred terms.

The list of MedDRA preferred terms used to identify any potential AESI is provided in *AESIs MeDDra.xlsx*.

All TEAEs identified using the search strategy described above will be medically evaluated during the study before the database lock and unblinding, by the sponsor to identify events which could possibly represent "remote spread of effect of toxin", or which are suggestive of "related to urinary incontinence or faecal incontinence" or "hypersensitivity reactions" due to study treatment administration. Events will be disregarded as an AESI if they are confounded by the presence of alternative clinical aetiologies (medical history, concomitant medication or diagnosis which could account for the symptoms); if they are considered to be local effects instead of distant spread, as judged by the site of injection; the time period between the last study treatment administration and event onset is not in accordance with the expected mechanism of action; or due to insufficient information/evidence to make an assessment.

In the summary data tabulations and listings, only the final list of AESIs confirmed by the sponsor as "a possible remote spread event", "hypersensitivity reactions" or "urinary incontinence and faecal incontinence" will be taken into account.

Besides the AESI listing, a summary table of AESI will be presented by SOC and PT (as described in Section 3.9) and include the following:

- Subjects with any AESI suggesting a possible remote spread of effect of the toxin,
- Subjects with any AESI related to urinary incontinence or faecal incontinence,
- Subjects with any AESI assessed as a potential hypersensitivity reaction.

3.9.3 Laboratory data

All laboratory data will be presented in the units of International System of Units (SI).

The following summaries will be provided:

- A summary of the actual and change from baseline in each laboratory parameter by treatment group and timepoint,
- A summary of the number and percentage of subjects experiencing Potentially Clinically Significant Abnormalities (PCSA). PCSA criteria for laboratory parameters are defined in appendix 2,

For PCSA table, the denominator should be the number of subjects with both a baseline and a post-baseline assessment at a given timepoint.

In addition, the following listings will be provided:

- A listing of all laboratory data. Out-of-reference-range values will be flagged as high (H), low (L), clinically significant [C]. Any unscheduled laboratory assessments will be flagged as such in the listings.
- Laboratory reference ranges,
- A listing will present all values of the parameter for a subject with at least a clinically significant abnormal value for hematology and biochemistry.
- A listing of PCSA. All data for a laboratory parameter will be displayed for a subject having at least one post-baseline PCSA (with flag indicating PCSA).
- A listing will present the urine drug screening and urine pregnancy test.
- A listing will be provided for subjects with abnormal result of drug abuse test.

Any clinically significant laboratory result abnormalities observed during the study will be reported as AEs.

3.9.4 Vital signs

Vital signs were recorded at the Screening Visit. A listing of vital sign data by treatment group, with unscheduled vital signs flagged.

Any clinically significant vital sign abnormalities observed during the study will be reported as AEs.

3.9.5 Physical examination

Physical examination was done at the Screening Visit. Date of exam and overall assessment will be listed and sorted by DB treatment group, and subject ID, with unscheduled physical examination flagged.
Any clinically significant physical examination finding observed during the study will be reported as AEs.

3.9.6 Antibodies

The following definitions will be considered:

- **Seroconverters** are subjects with ADA titer ≤1 for antibody at baseline and ADA titer > 1 post-treatment.
- **Seroreverters** are subjects which ADA titer > 1 at baseline and ADA titer <= 1 post-treatment.
- **Treatment boosted ADA** are subjects with ADA titer > 1 at baseline that were boosted to a higher-level following IMP administration (the ADA titer is greater than 100 times the baseline titer).
- **Incidence** is defined as the proportion of seroconverters or treatment-boosted subjects.

<u>The following conventions will be considered:</u>
- Percentage of seroconverters and incidence should be calculated using the number of subjects with a baseline and at least one ADA assessment post-treatment.
- Percentages of seroreverters and Treatment boosted ADA should be calculated using the number of subjects with ADA titer > 1 at baseline and at least one ADA assessment post-treatment.

The following summaries of ADA data will be provided:

- An overall summary of ADA status for binding and neutralizing with:
  - number and percentage of baseline results with titer > 1,
  - quartiles of the titer range,
  - number of seroreverters,
  - number of treatment-boosted,
  - number and percentage of seroconverters,
  - incidence.

A shift table from baseline to EOS/EW visit with the number and percentage of subjects separately for binding and neutralising ADA will be provided.

The following listings should be provided:

- A listing of any ADA data, including baseline BTX status (if the subject has previously received any treatment with BTX), and the status and titer for binding ADA and titer for neutralizing ADA,
- An overall ADA status of each subject,
- A listing for all subjects with binding ADA results with titer > 1 post baseline. This listing will include the combined information about baseline BTX status, primary efficacy results and TEAEs,
- A listing for all subjects with neutralizing ADA results with titer > 1 post baseline. This listing will include the combined information about baseline BTX status, primary efficacy results and TEAEs,

3.10 Missing data and other data issues

3.10.1 Missing data

For vestibular pain using a vaginal dilator, if a dilator size is not tested due to pain intensity, the pain will be scored at 10. If the test is not done due to other reasons, then the pain score will be considered as missing.

When calculating the composite score for the vaginal dilator induced pain and dilator size, there should be at least 6 pain scores (recorded by the investigator) to calculate the composite score. When missing, pain score will be imputed by the average of the previous (smaller dilator size) and the following (larger dilator size) valid values. For example, if the pain score for dilator size 4 is missing, then it will be imputed by the average of the pains scores for dilator size 3 and size 5 when both of them are available.

For mVPAQ, Pain Severity should have at least 4 (out of 6) valid inputs; Emotional Response should have at least 10 (out of 15) valid inputs; Cognitive Response should have at least 5 (out of 8) valid inputs; Life Interference should have at least 7 (out of 11) valid inputs; Sexual Function Interference should have at least 7 (out of 10) valid inputs; Self-Stimulation/Penetration Interference should have at least 3 (out of 5) valid inputs. The not applicable "N/A" entries are considered as valid inputs.

For mFSFI, Desire domain should have at least 2 (out of 2) valid inputs; Arousal and Lubrication domains should have at least 3 (out of 4) valid inputs; Orgasm, satisfaction and Pain domains should have at least 2 (out of 3) valid inputs. There should be at least 4 (out of 6) non missing domain scores to calculate the total score.

For PHQ-9, there should be at least 6 (out of 9) valid inputs to calculate the total score.

When calculating scores/subscale scores from questionnaires (e.g, mVPAQ mFSFI, and PHQ-9), the missing item value will be ignored, and the scores computed will be scaled to have the same possible maximum score as for the whole questionnaire (by dividing the possible maximum score based on questions with non-missing values and multiplying the possible maximum score for the whole questionnaire).

3.10.2 Missing or incomplete dates

In all listings, missing or incomplete dates should be left as they have been recorded. However, for calculation / sorting / assignation based on dates, the following methods will be used:

The most conservative approach will be systematically considered (i.e. if the onset date of an AE/concomitant medication is missing / incomplete, it is assumed to have occurred during the study treatment phase (i.e. a TEAE for AEs) except if the partial onset date or other data [stop date, …] indicates differently).

- A missing/incomplete date of medical history or disease diagnosis will be assumed to have occurred before any study treatment,
- A medication with partial start and stop dates will be considered as concomitant treatment, except if the partial dates indicate differently,
- If an AE or SAE onset date is partial or missing, the event will be allocated to the first treatment where onset could have occurred (taking into account date and time stopped),
- If the start date of a medication is partial or missing, the medication will be assigned to the most recent treatment received on or before the medication start date (taking into account date stopped).

Where this is possible, the derivations based on a partial date will be presented as superior inequalities (i.e.: for an AE started in FEB2004 after the administration performed on 31JAN2004, the days since last dose will be "≥2", similarly the duration of ongoing AEs or medication will be "≥xx" according to the start and last visit dates).

If the date of diagnosis for provoked vestibulodynia is partial, the duration of provoked vestibulodynia will be calculated as follow: 1) If only day is missing, only month and year will be used in calculation. 2) If month is missing, only year will be used in calculation. 3) No imputation should be done at the data level.

3.11 Derived data

The derived data are variables which are calculated from the raw data in the eCRF and not included in the database.

The following derived data will be calculated and included in the listings:

(1) Age

Subject age (years) will be derived as (screening year – birth year).

(2) BMI

BMI (kg/m^2) will be derived as Weight (kg)/[Height(cm)/100]^2.

(3) Changes from Baseline

Changes from Baseline will be calculated as a difference from Baseline (e.g. assessment at the visit – assessment at Baseline).

(4) Adverse event duration

The AE duration = AE end date - AE start date + 1. If the start and end dates of the adverse event are identical then "<1" day will be presented, otherwise, it will be calculated as (end date - start date)+1 and presented in days. If the recorded end date is CONT. (for continuing), the end date will be listed as "ongoing" and the duration will be approximated as "≥(last attended visit date – start date)+1" day(s). If the start date or the end date are partial the duration will be presented as a superior inequality "≥xx" day(s) [i.e.: ≥2 where start date=31JAN2004 and end date=FEB2004 or start date=JAN2004 and end date=01FEB2004].

(5) Concomitant therapy duration

The duration of concomitant treatments/physiotherapy etc. will be calculated as (end date - start date) +1. If the recorded end date is CONT. (for continuing) then the end date will be listed as "ongoing" and the duration will be approximated as "≥(last attended visit date – start date)+1" day(s). If the start date or the end date are partial, the duration will be presented as an inequality "≥xx" day(s) [i.e.: ≥2 where start date=31JAN2004 and end date=FEB2004 or start date=JAN2004 and end date=01FEB2004] but if both are partial or one is missing the duration will not be presented.

(6) Study day

Study day will be defined as '-1' for the day prior to first study drug administration and as '1' for the day of first study drug administration (i.e. day 0 does not exist).

(7) Study Baseline

Baseline is defined as the last value, or the derived value specified for the parameter, available prior to the first study IMP treatment administration (i.e. prior to the DB period). Baseline for the active dysport cycle analysis will be the same as Baseline for the DB period, i.e., Baseline of the first study cycle.

(8) Treatment cycle day

For events within the i^{th} cycle, the treatment cycle day = date of event - (date of (i)th injection) +1.

3.12 Rules and data formats

Data will be presented using an appropriate number of decimal places (i.e. the number of decimal places used does not imply undue precision). Raw data will be presented to the number of decimal places collected, and derived data will be presented to an appropriate number of decimal places. The appropriate number of decimal places will be determined by general practice, mathematical rationale or scientific rationale (e.g. age should be presented in whole numbers).

Mean, median, standard deviation and standard errors of the mean (SE) values will be reported to one decimal place greater than the raw/derived data that they summarize. Minimum and maximum values will be reported with the same precision as the raw data.

Percentages will be reported to one decimal place and 0% will not be presented. Percentages will be calculated using a denominator of all subjects in a specified population. The denominator will be specified in a footnote to the tables for clarification if necessary.

Percentiles (e.g., 25%, 75%) should be presented to one decimal place more than the raw/derived data.

Lower and upper confidence interval values should be presented to one decimal place more than the raw/derived data (i.e., to the same number of decimal places as the mean).

P-values will be reported to four decimal places (e.g.: p=0.0037), after rounding. P-values which are less than 0.0001 will be presented as '<0.0001'.

All values below or above a limit of detection (e.g. <0.1 or >100) will be listed as such.

All text fields must be left justified and numeric or numeric with some text specification (e.g.: not done, unknown, <4.5, …) must be decimal justified. Dates will be presented in the format [ddmmmyyyy] and times in the format [hh:mm].

3.13 Pooling of Centres

A subgroup analysis on individual or groups of centres will not be performed.

3.14 Covariates and analysis of subgroups

The following statistical analyses by pain onset subtype (randomization stratification factor) will be done:

- Demographics and baseline characteristics will be summarized by pain onset subtype, in the mITT population.
- Descriptive statistics for the change from Baseline to Week 6 and Week 12 in the vaginal dilator induced pain as reported on an 11-point NRS (using the

DMTS reported at Baseline) will be provided by pain onset subtype (primary and secondary PVD) on the mITT population for DB period.

- Descriptive statistics for the change from Baseline to Week 6 and Week 12 in the dilator size that provokes maximum tolerated pain will be provided by pain onset subtype (primary and secondary PVD) on the mITT population for DB period.
- An overall adverse events summary will be done; TEAEs and SAEs will be described by SOC and PT for DB period (please refer to Section 3.9 for more details on the summary strategy).

3.15 COVID-19 impact

The e-CRF was amended in order to collect data related to COVID-19 impact. These will be summarized as described thereafter.

Disposition

To assess the impact of COVID-19 pandemic on the trial, the following summaries will be provided:

- Proportion of subjects impacted and withdrawn due to COVID-19 pandemic.
    - the reasons of discontinuation related to COVID-19 will be presented separately.
    - the number of subjects exposed as well as impacted or infected will also be presented.
- Number of subjects with visits impacted by COVID-19 with the corresponding reason.
    - the number of subjects per visit as well as the number of subjects with a missing visit per timepoint.
    - the number of visits impacted due to COVID-19 with the reasons will be presented to assess the impact of COVID-19.

Demographic and Baseline characteristics

Summary statistics will be provided for demographic and Baseline characteristics, by DB treatment group in subjects recruited (based on informed consent date) before and after (inclusive) the start of COVID-19 pandemic [i.e. 27-Jan-2020 based on date when the Public health emergency was declared by Health and Human Services] in order to identify a potential change in the general subject's characteristics.

Exposure

To assess the impact of COVID-19 pandemic on the trial, the time elapsed between injections will be described for injections done before and after the start of COVID-19 pandemic (27-Jan-2020), which will be presented by cycle and one subject will be counted once per cycle and period. Any cycle that ongoing on 27-Jan-2020 or after 27-Jan-2020 will be considered as cycle after the start of COVID-19 pandemic.

4 COMPUTER SYSTEMS, SOFTWARE AND VALIDATION OF PROGRAMS

4.1 Software

All tables, listings and figures will be produced and statistical analysis performed using SAS version 9.4.

4.2 Validation programs

Medpace will provide a Validation Plan to Ipsen identifying the methods of validation.
Copies of the internal QC forms produced for the validation process and Medpace's sign-off forms will be provided to the sponsor to support the validation.

4.3 Restitution of the programs

All programs (including Macros and analysis datasets) producing the tables, listings and statistical output along with associated logs should be given to the sponsor when the tables, listings, figures and statistical analysis has been finalised.

5 CHANGES FROM PROTOCOL

5.1.1 Schedule of assessments

The descriptive analysis to evaluate/describe any possible centre effect in protocol (section 11.4.4) is removed due to the small sample size.

5.1.2 Data Analyses Timepoints

The intermediate and primary analyses will not be performed. All analyses related to Stage 2 will not be performed since the study was terminated before Stage 2 begins. Please refer to Section 1.3.5 for more details.

5.1.3 COVID-19 Pandemic Start Date

The date 27-Jan-2020, when public health emergency was declared by Health and Human Services in the US, is used instead of the date specified in the protocol (i.e., when first subject randomised in Cohort 6) in order to describe any possible impact of COVID-19 pandemic on subjects followed-up in all cohorts, including during the open-label cycles.

5.1.4 Per Protocol (PP) population

No Per Protocol population will be defined as it was planned for the analysis of Stage 2 and the study was terminated before Stage 2 begins. Stage 1 efficacy analyses consist in descriptive only on the mITT population.

6 REFERENCES

1. Rosen, C. Brown, J. Heiman, S. Leiblum, C. Meston, R. Shabsigh, D. Ferguson, R. D'Agostino, R. The Female Sexual Function Index (FSFI): a multidimensional self-report instrument for the assessment of female sexual function. Journal of sex & marital therapy 2000; 26(2): 191-208.

7 APPENDICES

7.1 Appendix 1: List of PCSA criteria

PCSA for blood chemistry parameters:

Parameter	PCSA
ALT/SGPT	≥ 3 x ULN U/L
AST/SGOT	≥ 3 x ULN U/L
GGT	≥ 3 x ULN U/L
ALP (Alkaline Phosphatase)	≥ 1.5 x ULN U/L or ≥ 3 x ULN U/L
Creatinine	≥ 150 µmol/L (Adults)
Blood Urea Nitrogen	> 10 mmol/L
Chloride	≤ 90 mmol/L ≥ 115 mmol/L
Calcium	<1.75 mmol/L >3 mmol/L
Sodium	<130 mmol/L > 150 mmol/L
Potassium	<2.5 mmol/L >6 mmol/L
Total cholesterol	> 6.22 mmol/L
Triglycerides	> 2.26 mmol/L
HbA1c	> 8%
Albumin	≤ 25 g/L
Direct bilirubin	> 2 ULN µmol/L
Total bilirubin	> 2 ULN µmol/L
Phosphate	< 0.75 mmol/L > 1.8 mmol/L
Total protein	< 40 g/L

PCSA for haematology parameters:

Parameter	PCSA
White blood cell	≤ 2.8 10^9/L ≥ 16 10^9/L
Red blood cell	< 3 10^12/L > 6 10^12/L
Hematocrit	≤ 0.32 and 0.03 decrease from baseline L/L (female) ≤ 0.37 and 0.03 decrease from baseline L/L (male)
Lymphocytes	< 0.3 10^9/L > 11.0 10^9/L
Neutrophils	< 1 10^9/L > 13 10^9/L
Monocytes	> 2.3 10^9/L
Basophils	> 10^9/L
Eosinophils	> 10^9/L
Hemoglobin	≤ 95 g/l (female) ≤ 115 g/l (male)
Platelet count	<75.1 10^9/L ≥ 700 10^9/L