Document Name: dfr52120237-stat-plan

Clinical	PPD 7:26:44 GMT+0000

Clinical	PPD 019 13:53:19 GMT+0000

Clinical	PPD 14:20:58 GMT+0000

Approved

STATISTICAL ANALYSIS PLAN

PROTOCOL TITLE: A MULTIPLE-DOSE, DOUBLE-BLIND, RANDOMISED, PLACEBO-CONTROLLED STUDY TO EVALUATE THE EFFICACY AND SAFETY OF DYSPORT FOR THE TREATMENT OF PAIN ASSOCIATED WITH HALLUX ABDUCTO VALGUS
D-FR-52120-237

This statistical analysis plan is based on:
PROTOCOL VERSION AND DATE: FINAL VERSION 6.0 – 03 JULY 2019

SAP Version	Date
Final Version 1.0	20 September 2019

APPROVAL PAGE

STUDY NUMBER:	D-FR-52120-237
PROTOCOL TITLE:	A MULTIPLE-DOSE, DOUBLE-BLIND, RANDOMISED, PLACEBO-CONTROLLED STUDY TO EVALUATE THE EFFICACY AND SAFETY OF DYSPORT FOR THE TREATMENT OF PAIN ASSOCIATED WITH HALLUX ABDUCTO VALGUS
SAP VERSION:	Final Version 1.0
SAP DATE:	20 SEP 2019

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		
PPD	IPSEN	PPD		
PPD	Medpace	PPD		
PPD	Medpace	PPD	PPD	

HISTORY OF CHANGES

Version Number	Date	Description/Rational for change

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

AE	Adverse event
AESI	Adverse Event of Special Interest
ANCOVA	Analysis of Covariance
BTX	Botulinum Toxin
BTX-A	Botulinum Toxin Type A
BTX-A-HAC	Botulinum Toxin Type A Hemagglutinin Complex
CMC	Chemistry Manufacturing Control
CRO	Contract research organisation
CSR	Clinical Study Report
DB	Double-blind
eCRF	Electronic case report form
FDA	Food and Drug Administration
FFI	Foot Function Index
GCP	Good Clinical Practice
HV	Hallux Valgus (angle)
IA	Interim Analysis
IB	Investigator's brochure
ICH	International Conference on Harmonisation
IMP	Investigational Medicinal Product
IRB	Institutional review board
IRT	Interactive response technology
ITT	Intent-to-treat
MAR	Missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed Model for Repeated Measures
mFFI	Modified Foot Function Index
NPRS	Numeric Pain Rating Scale
OL	Open-label
PCSA	Potentially Clinically Significant Abnormalities
PGI-I	Patient Global Impression of Improvement
PGI-S	Patient Global Impression of Severity
PP	Per protocol
SAE	Serious adverse event

SAP	Statistical analysis plan
SD	Standard deviation
SOP	Standard Operating Procedure
SUSARs	Suspected Unexpected Serious Adverse Reactions
TEAE	Treatment emergent adverse event
US(A)	United States (of America)
WHO	World Health Organisation

1 INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to outline the planned analyses to be completed to support the completion of the Clinical Study Report (CSR) for protocol D-FR-52120-237 Amendment 5. It describes the rules and conventions to be used in the analysis and presentation of data, the data to be summarized and analyzed, including specificities of the statistical analyses to be performed.

Exploratory analyses not necessarily identified in this SAP may be performed to support the clinical development program. Any post-hoc, or unplanned, analyses not identified in this SAP will be clearly identified in the respective CSR.

An interim analysis (IA) will be conducted based on the first 110 subjects who have been followed up at least 12 weeks in the study. The interim analysis has been powered to detect futility and efficacy. More information regarding the IA can be found in Section 3.2.1 and the Interim Analysis Charter.

The SAP is to be finalized prior to the interim analysis. The corresponding tables, figures and listings shells will be provided separately.

Any deviations from the SAP after database lock will be documented in the CSR (section 9.8 “Changes in the conduct of the study or planned analyses” as per ICH E3).

2 PROTOCOL OVERVIEW

2.1 Study Objectives and Hypotheses

2.1.1 Primary objective

The primary objective of this study is to assess reduction in pain in adult subjects with hallux abducto valgus (HV) with Dysport (300 U and 500 U) as compared with placebo using a Numeric Pain Rating Scale (NPRS).

2.1.2 Secondary objectives

The secondary objectives of the study are:

- To assess functional improvement in Dysport (300 U and 500 U) as compared with placebo using the modified Foot Function Index (mFFI) disability subscale.
- To assess reduction in foot pain in Dysport (300 U and 500 U) as compared with placebo using the mFFI pain subscale.
- To assess improvement in activity limitation associated with foot pain in Dysport (300 U and 500 U) as compared with placebo using the mFFI activity limitation subscale.
- To evaluate the quality of life using the 36-item Short Form (SF-36).
- To evaluate angular displacement of the hallux using radiographs.
- To evaluate the patient’s global impression of improvement and severity associated with foot pain and disability.
- To evaluate the clinical safety and efficacy of Dysport following repeated treatment cycles.

2.1.3 Exploratory objectives

The exploratory objective of this study is:

- CCI [redacted]

2.1.4 Primary efficacy hypotheses

The primary efficacy hypotheses are as follows:

- H0: There is no difference between treatment with Dysport 300 U and Dysport 500 U versus treatment with Placebo with respect to the change from baseline in the

mean daily NPRS score averaged for the 7 consecutive days prior the double-blind (DB) Week 8 visit.

- H1: There is a difference between treatment with Dysport 300 U or Dysport 500 U and treatment with Placebo with respect to the change from baseline in the mean daily NPRS score averaged for the 7 consecutive days prior the DB Week 8 visit.

2.2 Overall Study Design and Investigational Plan

This is a multiple-dose, randomized, parallel-group, DB, placebo-controlled study to evaluate the efficacy and safety of Dysport at doses of 300 U and 500 U compared with placebo in adult subjects suffering from clinically significant HV who have not undergone surgery for their condition. The study will consist of a screening period, followed by a DB period where subjects will receive a single treatment with either Dysport (300 U and 500 U dose groups) or placebo (Cycle 1) followed by an open-label (OL) extension period where subjects may receive up to two additional cycles of Dysport (300-500U) in the affected foot. As such, the maximum duration of participation in the study for a given subject will be 36 weeks. The randomization will be stratified by patients having either unilateral or bilateral HV.

2.3 Sample Size Determination and Power

Several lines of evidence have established that a reduction of approximately 2 points or 10% to 30% on the NPRS represents a clinically important difference for evaluations of pain intensity in numerous musculoskeletal disorders, as well as specifically in HV patients following treatment. Several interventional studies have demonstrated treatment-placebo differences in the NPRS ranging from 1.5 to 2.0 points in HV patients. Based on these findings, a difference of 1.5 points between Dysport and Placebo is anticipated in the change from baseline in the mean daily NPRS score averaged for the 7 consecutive days prior to the Week 8 DB visit.

A sample size of 165 subjects (55 subjects per treatment group) is required to demonstrate the superiority of each of the two Dysport doses (300 U and 500 U) over placebo. This calculation is based on the following assumptions:

- a mean difference of 1.5 points observed between the Dysport and placebo groups in the mean daily NPRS score averaged for the 7 consecutive days prior to the Week 8 visit in the DB period,
- a common standard deviation of 2.5,
- a treatment group ratio of 1:1:1,
- a Hochberg procedure is implemented to control the family-wise Type 1 error rate at one-sided 2.5% level for comparisons of two Dysport doses versus placebo,
- a minimal power of 80%, corresponding for a comparison based on power using a Type I error rate from a one-sided test at 1.25% level.

An interim analysis will be conducted after the first 110 randomised subjects have been followed up for at least 12 weeks. The interim analysis is described in section 3.2.1. This interim analysis plan results in a minimal power of at least 80% to detect a treatment effect of 1.5 units in a Dysport arm.

2.4 Randomization and Blinding

All Investigational Medicinal Product (IMP) will be similar in size, colour, smell, taste and appearance allowing the blinded conditions of the study to be maintained.

Subjects and investigators will remain blinded to treatment assignment during the Double Blind period. The sponsor's randomization manager who is a statistician independent from the study will prepare:

- A list of randomization numbers (List A). It will be produced in blocks, on a balanced ratio (1 placebo: 1 Dysport 300 U: 1 Dysport 500 U) and will be stratified by unilateral and bilateral HV.
- A list of treatment numbers/treatment which will be dispatched to the sites (list B). It will be produced in blocks, on a balanced ratio (1 placebo: 1 Dysport 300 U: 1 Dysport 500 U).

2.5 Schedule of Assessments

Schedule of assessments is presented in section 5.1 from the protocol.

3 PLANNED ANALYSES

3.1 Safety Monitoring

No independent DMC will be used in this study.

3.2 Interim Analysis / Primary Analysis

3.2.1 Interim Analysis

An interim analysis will be conducted after the first 110 randomised subjects have been followed up for at least 12 weeks. The aim of this interim analysis is to both assess futility and the potential for early stopping due to efficacy of one of the Dysport group as compared to the Placebo group. A decision to continue with the study will be determined by an independent DMC based on the outcome of the interim analysis. No interruption to recruitment will occur whilst the decision-making process is ongoing. The decision rule is described in section 5.7.1.2.

Details regarding operational aspects of the interim analysis are described in the Interim Analysis Charter.

3.2.2 Primary Analysis

A primary analysis will be conducted once all subjects have completed the Week 12 of the DB period. At this point, all subjects will have completed the Week 8 visit and had adequate safety follow-up.

3.3 Final Analysis

A final analysis will be carried out once all subjects have completed end of study evaluation, after the database lock.

4 ANALYSIS POPULATIONS

4.1 Safety population

All subjects who received at least one dose of IMP administration (including only partial administration). Subjects will be analyzed using subjects as treated for the safety analyses.

4.2 Intention-To-Treat (ITT) population

All randomized subjects (i.e. who were randomly allocated to a treatment group by IRT). Subjects will be analyzed using subjects as randomized for the efficacy analyses.

4.3 Per Protocol (PP) population

All subjects in the ITT population for whom no major protocol deviations (which may interfere with efficacy evaluation) occurred until Week 8 of the DB period.
Major protocol deviations are listed in the protocol deviation plan.
Data will be excluded from the Per Protocol population on a subject basis and not on a visit basis.

4.4 Active Treatment population

All randomized subjects who received at least one dose of Dysport (including only partial administration) during the DB or OL period. Subjects will be analyzed using subjects as randomized for the efficacy analyses and using subjects as treated for the safety analyses.

4.5 Mis-Stratification

For statistical analysis, in the event a subject is stratified incorrectly, "randomized stratum" will be used rather than "actual stratum". Preservation of the initial randomization in analysis is important in preventing bias and in providing a secure foundation for statistical tests (ICH E9).

5 STATISTICAL METHODS/ANALYSES

The statistical analyses will be performed by Medpace in accordance with ICH E9 guideline and guidelines presented in section 8.

Overall, the analysis strategy is to evaluate efficacy and safety of two doses 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.
Statistical inferences will only be performed on the DB period until Week 12.

5.1 General Considerations

All statistical analyses will be performed using the SAS® software version 9.4.

5.1.1 Outputs Presentation

5.1.1.1 Tables Header

Depending on the type of data, the summary tables will be presented as follows:
- For disposition, demographic data and baseline data description: by treatment group and overall,
- For efficacy and safety data: by treatment group only.

5.1.1.2 Presentation of Treatment Group

Double-Blind Period Analyses

For the analysis of the DB period, tables will be displayed using the following treatment group labels, in the following order:

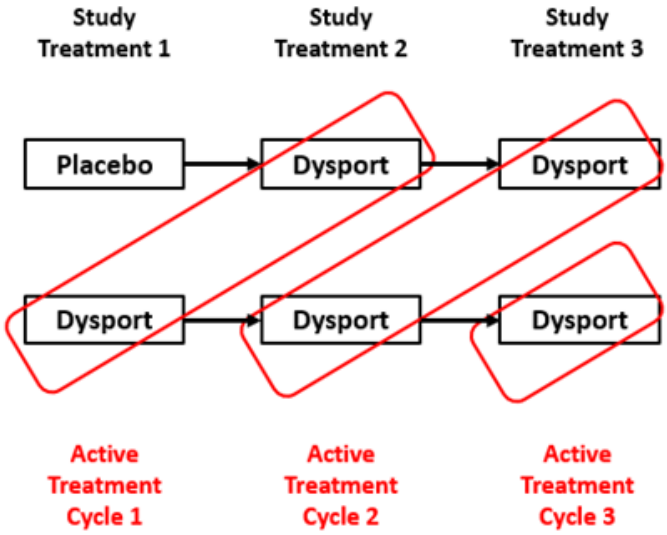
- Placebo
- Dysport 300 U
- Dysport 500 U
- All Dysport

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 and 3) active treatment cycles. In each treatment cycle, the summary will be by whether the first active treatment is in DB or OL (when applicable), and in total.
- Summary by sequence of doses and active treatment cycle: these tables will present the efficacy of the first two doses for subjects having at least two doses, and the three doses for subjects having three doses.
- Summary by dose received at each active treatment cycle: to evaluate the safety of each dose at a given cycle.

The type of summary table(s) used for each endpoint are described in sections 5.7 and 5.8.

5.1.2 Descriptive Statistics

All raw and derived variables will be listed and described using summary statistics. For categorical variables, summary statistics will be displayed using descriptive statistics by frequency count and percentages by category. The missing category will be presented if there is one missing category for at least one 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.

5.1.3 Baseline value

Unless otherwise specified, baseline is defined as the last non-missing measurement taken prior to first IMP administration (including unscheduled assessments).

5.1.4 Reference Start Date and Study Day

Reference start date is defined as the day of the first IMP administration. For subjects who are randomized but do not receive study drug, randomization date should be taken as reference start date for efficacy analysis (ITT).
The day of the first IMP administration will be Day 1. Study day will be calculated as:
- The difference between the event date and the reference date plus one day, if the event is on or after the reference date.
- The difference between the event date and the reference date, if the date of event is prior to the reference date.

Study day will appear in any listings where an assessment date or event date appears. In case of partial or missing event date, study day will appear missing while any associated durations will be presented based on the imputations described in appendix A3.

5.2 Randomization, Disposition and Population

A listing presenting randomization details will be provided.
A listing of re-enrolled patients will be provided.
Following disposition summaries and listings will be provided:

- Summary table with the number and percentages of randomized subjects per site by DB treatment group on the ITT population,
- Summary table with the number and percentage of subjects screened, screen failed, reason for screen failure, randomized, completed, withdrawn and reason for withdrawal by DB treatment group on all screened subjects,
- Summary table with the number and percentage of subjects withdrawn and reason for withdrawal by Active Treatment Cycle on the Active Treatment population.
- Summary table on duration of subject participation in the study on the ITT population. The definition of the duration of subject participation is date of consent – date of the last study visit + 1.
- Listing of dates of visit on the ITT population,
- Listing of screen failure subjects on all subjects,

- Listing of withdrawn subjects on the ITT population.

Following analysis population summaries and listings will be provided for the ITT population:
- Listing of subjects violated inclusion criteria,
- Listing of subjects fulfilled exclusion criteria,
- Summary of the number and percentage of subjects in each analysis population by DB treatment group and overall, based on all randomized subjects with reasons for exclusion from each analysis population,
- Listing including flag for each analysis population and reason for exclusion from each population.

5.3 Protocol Deviations

A list of major protocol deviations and any action taken have been defined in the Protocol Deviation Plan. Patients with major protocol deviations will be determined before unblinding of the study. A final review will be done during the blind data review and documented in a separate document and the impact on the per protocol population will be assessed. Confirmation of major deviations related to treatment and dose will be done after unblinding.
Number and percentage of subjects with protocol deviations will be summarized by deviation category for all randomized subjects. A Listing of major protocol deviations will also be provided.

5.4 Demography and Other baseline characteristics

All demographic and baseline characteristics summaries will be provided for the ITT and safety population (if different from ITT population). Disease characteristics (HV bilaterality, baseline severity [average NPRS score], screening and baseline Hallux Valgus (HV) angle, screening and baseline intermetatarsal angle, use of insoles, time since using insoles, type of insole, whether a device for ePRO is provided) will be included. No statistical comparison between treatment groups will be performed.
Listings will also be provided for all the summaries listed above.

5.5 Medical history, non-drug therapies, medications and surgical procedures

Medical and surgical history, non-drug therapies and surgical procedures will be coded using the latest version of MedDRA in effect within IPSEN at the time of database lock. Medications will be coded using the latest version of WHO-Drug dictionary in effect within IPSEN at the time of database lock.

Medication, non-drug 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 only, Prior and Concomitant, or Concomitant only:

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, and surgical procedures will include "P" only, summary tables on concomitant medications, non-drug therapies, and surgical procedures will include "C" and "PC".

See detailed rules in appendix A3 for classification of prior and concomitant medication/non-drug therapies, surgical procedures in case of partial/missing date.

The therapeutic class will correspond to the second level of ATC code, that is, corresponding to the first 3 figures.

The following summaries, presenting count and percentages of subjects will be provided for the ITT population:

- Medical and surgical history by primary system organ class (SOC) and preferred term (PT),
- Prior medications (P) by ATC class and PN (ATC level 2),
- Concomitant medications (PC, C) by ATC class and PN (ATC level 2),
- Prior non-drug therapies (P) by primary SOC and PT,
- Concomitant non-drug therapies (PC, C) by primary SOC and PT, and
- Concomitant surgical procedures (C) by primary SOC and PT.

Listings will be provided for all the summaries listed above. These listings should include a flag indicating the category (P, PC, C) as described in the table above.

5.6 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.

5.7 Efficacy

5.7.1 General Considerations

5.7.1.1 Significance Testing and Estimations

For the primary efficacy analysis, a Hochberg procedure will be applied to control the global type I error at one-sided 2.5% significance level.
For other efficacy analyses, all statistical tests will be performed one-sided with a type I error rate set at 2.5%.
For safety analyses, no formal statistical analyses will be carried out; only descriptive statistics will be provided.

Descriptive statistics will be provided for:

- the DB period, by DB treatment group for the ITT population.
- the Active Treatment Cycle, on the Active Treatment population either by number of active treatment cycles (all doses combined) or by sequence of dose in active treatment cycles (see section 5.1). The type of analysis will be specified for each endpoint.

5.7.1.2 Statistical/analytical issues

Adjustments for Covariates

The randomization is stratified by unilateral and bilateral HV. Therefore, the planned statistical analysis for the demonstration of the superiority of Dysport compared to Placebo will include the stratification parameter (unilateral or bilateral HV) in the model as a factor. An analysis without the stratification parameter will also be performed to evaluate the effect of the stratification parameter on the results.

Handling of Dropouts or Missing Data

The strategy for missing data handling is specified in each relevant section.

Interim Analyses and Data Monitoring

Please refer to section 3.

Multicentre Studies

This is a multicentre study conducted in US. The centre effect will be explored using analysis centre. Analysis centre will be generated using the following procedure:

- Centres are sorted by the total number of ITT subjects and centre ID in ascending order.
- If the first (smallest) centre has less than 6 ITT subjects, it will be pooled with the next centre on the sorted list (i.e., the smallest of other sites) until the pooled (analysis) centre has at least 6 ITT subjects.
- The previous step is repeated for the rest of the centres until all analysis centres have at least 6 ITT subjects. If the total number of ITT subjects in the rest of the centre(s) is less than 6, they will then be included in the previous analysis centre.

The primary analysis will be repeated with the analysis centre as an additional factor in the model. If the hypothesis that the centre effect does not exist is rejected, the analysis centre will be included in all the analyses.

Multiple Comparisons/Multiplicity

Due to the multiple comparisons, significance levels will be adjusted using the Hochberg adjustment to reduce the risk of false positive results. The raw p-values will be included in the summary tables.

To account for the interim analysis, the overall type I error for each comparison is controlled at the one-sided 0.025 level using O'Brien Fleming spending per Lan-DeMets spending function specification. Treatment comparison with an O'Brien Fleming spending corresponds to a nominal one-sided alpha of 0.0062 at the interim and 0.0231 at the final analysis.

To implement the Hochberg procedure, decision rules are the following:

- At interim look:
  Compare larger p-value to 0.0062
  - If <larger of 2 p-values> <0.0062 then stop and declare evidence of effect for each arm,
  - If not, compare <smaller of 2 p-values> to 0.0062/2. If <smaller of 2 p-values> <0.0062/2, then conclude evidence of effect for arm with the smallest p-value

- At final look (if outcome of IA is to continue):
  Compare larger p-value to 0.0231
  - If <larger of 2 p-values> <0.0231 then stop and declare evidence of effect for each arm
  - If not, compare <smaller of 2 p-values> to 0.0231/2. If <smaller of 2 p-values> <0.0231/2, then conclude evidence of effect for arm with the smallest p-value

The non-binding futility boundary is set to declare futility if the one-sided p-value for a comparison is above 0.30.

During the interim analysis, if futility is concluded for the primary endpoint, the secondary endpoint mFFI pain subscale score (change from baseline in the daily mFFI pain subscale score averaged over the 7 consecutive days prior to the Week 8 DB visit) will be analysed using MMRM as described in Table 3. If a one-sided p-value is ≤ 0.05, then declare evidence of trend for the corresponding arm; Otherwise we conclude not having evidence of trend for the corresponding arm.

Examination of Subgroups
Subgroup analyses will be detailed in the concerned sections.

5.7.2 Analysis of Primary Efficacy Endpoint

5.7.2.1 Endpoint, Treatment Effect and Estimand Definition

The primary efficacy endpoint is the change from baseline in foot pain, as measured by the daily NPRS score averaged over the 7 consecutive days prior to the Week 8 DB visit. Baseline is defined as the daily NPRS score averaged over the 7 consecutive days prior to the baseline visit.
To assess the primary objective, the hypothetical estimand is defined by the following key attributes:

- Population: Subjects in the ITT.
- Variable: Change from baseline in foot pain regardless of whether the subject has an intercurrent event during the study.
- The potential intercurrent events are:
  (1) Use of prohibited pain medication,
  (2) Surgery for HV.
- Population-level summary: Change from baseline in foot pain (estimate of the treatment effect).

No treatment switch or interruption is possible as the primary efficacy endpoint is evaluated following a single injection of study treatment. Intercurrent event (1) is unlikely to occur, as the evaluation period for the primary efficacy endpoint is up to Week 8 (short period). Intercurrent event (2) is also unlikely to occur. All subjects with this level of pain severity and angular displacement are considered surgical candidates theoretically. However, the upper limit on HV angle (30 degrees) will likely limit the need of a "rescue" surgery.
Subjects are expected to continue follow-up assessments regardless of these two intercurrent events. NPRS scores will be used as observed.

5.7.2.2 Primary Analysis

Descriptive statistics (including the summary of raw and change from baseline by treatment group and timepoint) will be provided for the DB period, by DB treatment group for the ITT population. Average score will be calculated if there are at least 4-days of e-diary completed, otherwise, the average score will be considered missing.
A mixed model for repeated measures (MMRM) on change from baseline in foot pain as measured by the daily NPRS score averaged over the 7 consecutive days prior to each scheduled assessment timepoint of the DB period (up to Week 12) will be used to evaluate the estimand and compare treatment groups. Missing data will be considered missing at random (MAR). For subjects without any available post-baseline average scores, the Week 8 average pain score will be imputed as follows:

- If a subject discontinues because of lack of efficacy, the Week 8 average pain score will be imputed by the mean of placebo group at Week 8,
- If a subject discontinues for other reason that lack of efficacy, the Week 8 average pain score will be imputed by the mean of the subject's treatment group at Week 8,
- If a subject completes the study, but has no valid diary (less than 4 diary days completed) post-baseline, the Week 8 average pain score will be imputed by the mean of the subject's treatment group at Week 8.

This model will include the fixed categorical effects of treatment group, visit, treatment group-by-visit interaction, and the stratification parameter as fixed categorical covariates and the baseline value as fixed continuous covariate. The treatment group factor will have three levels (Dysport 300 U, Dysport 500 U and placebo), the factor visit three levels (Week 4, Week 8 and Week 12) and the stratification parameter two levels (unilateral and bilateral HV).

The following estimates will be provided:
For each of the 3 treatment groups, the least squares mean of the change from Baseline to Week 8 in foot pain and associated 2-sided 95% confidence.
For each of the 2 Dysport groups, the least squares mean of the difference against placebo with a 2-sided 95% confidence interval and p-value.
See Appendix A1 for the SAS code.

The study will be considered successful if the superiority of at least 1 Dysport group relative to placebo is demonstrated.
Non-parametric tests may also be carried out to evaluate robustness of the results if the linearity, normality, homoscedasticity assumptions are violated.

The daily pain scores will be listed along with the 7-day average prior to the scheduled visits. For any subjects and visits where the 7-day average was not able to be calculated, the imputed value will be provided and flagged in the listing.

5.7.2.3 Sensitivity Analysis

Three sensitivity analyses will be performed to investigate the robustness of the primary efficacy analysis. All of them will only impute the missed Week 8 data if neither Week 4 or Week 12 value is available. For multiple imputation, missing data will be imputed 100 times to generate 100 complete data sets using the regression

method. For each of the 100 imputed datasets, the corresponding endpoint will be constructed, and the estimates from the 100 fitted models will be combined. Random seed = 3575 will be specified for the multiple imputation.

(1) The primary efficacy analysis will be re-run imputing missing data as described in Table 1~~Table 1~~.

Table 1 Rules for Missing Data Associated with the NPRS

Reason for missing data	Imputation of the missing NPRS score
Subject completed partially (≤3 days) the 7-day NPRS e-diary	Multiple imputation based on subject with similar characteristics on the same treatment group. Linear regression will be used with gender, stratification parameter as factors, baseline average NPRS score and age as covariate.
Subject withdrawal before Week 8 of the DB period due to lack of efficacy	NPRS score imputed as the mean of placebo value at Week 8
Subject withdrawal before Week 8 of the DB period due to other reason	Multiple imputation based on subject with similar characteristics in the same treatment group
Subjects not withdrawn before Week 8 of the DB period with a missing NPRS at Week 8	Multiple imputation based on subject with similar characteristics within the same treatment group

DB=double blind; e-diary=electronic diary; NPRS=numeric pain rating scale

(2) If there are more than 5% subjects without any post Baseline data, a tipping point analysis will be performed as described below.

Tipping Point Analysis

Multiple imputation method based on the missing at random assumption will be used for each treatment group separately. Random seed = 3565 will be specified. Variables included in the imputation model are age, gender, stratification parameter (unilateral and bilateral HV), and baseline average NPRS score. In each treatment group, a penalty δ will be added to the imputed values (δ = 0 corresponds to the missing at random assumption). The smallest value of δ that reverses the conclusion for primary analysis will be provided.

(3) Because it was noticed that several subjects had a pain score of 0 in between days of very high pain, the primary analysis will be repeated using the median NPRS pain score in the 7 consecutive days prior to each visit. It is possible that 0 was a default if the ePRO NPRS entry was started but not saved.

If there are at least 4 values available for Week 4, Week 8 or Week 12 (within 7 days prior to each visit), the data available will be used to determine the median NPRS pain score at that visit and no imputation will be needed. Otherwise, the Week 8 median score will be imputed using the same imputation method for primary analysis.

5.7.2.4 Supplementary Analysis

Supplementary analyses will be performed in order to complement the primary estimand. The primary efficacy analysis will be first re-run on the PP population. In addition, three other estimands will be considered:

(1) Proportion of responders in the NPRS pain score at Week 8 DB. A responder is defined as a subject with 20% decrease from baseline (30%, 40% and 50% cut-off will also be considered).

The NPRS pain score is the daily score averaged over the 7 consecutive days prior to the visit as described in the primary analysis. If the Week 8 average score is not available and average score for the next scheduled visit (Week 12) is available, the Week 8 average score will be imputed with the average score for the next scheduled visit (Week 12). If no available average score at Week 8 nor the scheduled visit after that (Week 12), the subject will be considered as a non-responder.
A logistic regression model will be used to estimate the proportion of responders within each treatment group. The model will include the treatment group and the stratification parameter as fixed categorical covariates and the baseline value as a continuous covariate.
See Appendix A1 for the SAS code.

(2) The mean change in the area under the curve (AUC) of the daily NPRS score averaged over the 7 consecutive days prior to each visit up to Week 8 of the DB period.

The AUC is calculated as $(a_1 + 2a_2 + 2a_3 + 2a_4 + 2a_5 + 2a_6 + a_7)/2$, where $a_1, a_2, \dots, a_7$ are the daily scores of the 7 consecutive days prior to each visit. Missing values will be imputed as follows:

- If missing input(s) at the beginning (e.g. first several days) of the 7 consecutive days, it (they) will be imputed with the first available input within these 7 days. (e.g., if 1st and 2nd days missing, 3rd day's input available, then 1st and 2nd days' inputs will be imputed with the 3rd day's value.)
- If missing input(s) at the end of the 7 consecutive days, it (they) will be imputed with the last available input within these 7 days.
- If missing in the middle, it will be imputed with the average of the first available inputs before and after the missed day. (e.g., if 1st, 4th days values available and 2nd , 3rd days missing, then 2nd, 3rd days' values will be imputed with the average of 1st and 4th days' values.)
- If there is no input within the 7 consecutive days, the imputation method for the primary analysis will be used.

Analysis of Covariance (ANCOVA) model on the change from baseline in AUC prior to each scheduled assessment timepoint of the DB period (up to Week 8) will be used to evaluate the estimand and compare treatment groups.
This model will include treatment group and the stratification parameter as fixed categorical covariates and the baseline value as a continuous covariate.

(3) Repeat the primary analysis using worst NPRS pain score in the 7 consecutive days prior to each visit instead of the average.

If there are at least 4 values available for Week 4, Week 8 or Week 12 (within 7 days prior to each visit), the data available will be used to determine the worst NPRS pain score at that visit and no imputation will be needed. Otherwise, the Week 8 worst score will be imputed using the same imputation method for primary analysis.

5.7.2.5 Subgroup Analysis

Descriptive statistics for the primary efficacy endpoint will be provided on the stratification parameter (unilateral or bilateral HV) on the ITT population by DB treatment group.

5.7.3 Analysis of Secondary Efficacy Endpoints

Subjects are expected to continue follow-up assessments regardless of the intercurrent events define above. Scores will be used as observed. No imputation of missing data will be done.

According to the scale and as detailed in the following section, an MMRM model or a mixed-linear-generalized model on change from baseline including each scheduled timepoint up to Week 12 DB will be used to evaluate the estimands and compare treatment groups. Missing data will be considered as MAR.

Only descriptive statistics will be used to present the estimands after Week 12 DB visit, including the active treatment cycles.

5.7.3.1 Foot pain NPRS score

The following summaries and analysis output will also be provided for the foot pain NPRS score.

Table 2 Foot Pain NPRS Score - Endpoints and Corresponding Analyses

Endpoint	Analyses
The change from baseline in the daily NPRS score averaged over the 7 consecutive days prior to scheduled visits.	- descriptive statistics: actual values and change from baseline values will be presented at all timepoints including Week 8. - Frequency table will be presented for number of days completed by treatment group. DB: - MMRM as described in the primary efficacy analysis. The same outputs will be provided for Week 4 and Week 12. Active Treatment Cycle: - A summary by number of active treatment cycles - A summary by sequence of active doses of each treatment cycle.

5.7.3.2 mFFI

The mFFI (provided in protocol Appendix 1) consists of a total of 21 items grouped into three subscales: pain (seven questions), disability (nine questions), and activity limitation (five questions). The mFFI items are rated using numeric rating scales ranging from 0 to 10 and cover a period of the 'past' 24 hours. The poles are labelled "no pain" and "worst pain imaginable" (pain), "no difficulty" and "so difficult unable to do" (disability), and "none of the time" and "all of the time" (limitations). For each item, the subject is asked to record the number value which best corresponds to the effect of the foot complaints.

To obtain a subscale score, the item scores for a given subscale (i.e. pain, disability or activity limitation subscales) are totalled and divided by the maximum total possible and then multiplied by 100. Each subscale score, as well as the total score, will range from 0 to 100. If a patient did not perform the task listed in the question, they will be instructed to mark the item as not applicable (N/A).

Subscale and total scores will be calculated when at least 50% of the items for that subscale/total score are present (not including the N/A answers). The average will be calculated only if there is at least 4 out of 7 days completed.

Table 3 mFFI - Endpoints and Corresponding Analyses

Endpoint	Analyses
The change from baseline in the daily mFFI disability subscale score averaged over the 7 consecutive days prior to each post-treatment visit.	- Descriptive statistics: actual values and change from baseline values will be presented at all timepoints. DB: - MMRM at Week 4, Week 8 and Week 12. MMRM setup is similar to the primary analysis. - The MMRM above will be repeated using the median instead of mean - Frequency table will be presented for number of days total score calculated by treatment group. - Descriptive statistics: score for each item by treatment group. Active Treatment Cycle: - A summary by number of active treatment cycles
The change from baseline in the daily mFFI pain subscale score averaged over the 7 consecutive days prior to each post-treatment visit.	
The change from baseline in the daily mFFI total score averaged over the 7 consecutive days prior to each post-treatment visit.	
The change from baseline in the daily mFFI activity limitation subscale score averaged over the 7 consecutive days prior to each post-treatment visit.	

5.7.3.3 Patient Global Impression (PGI)

Patient Global Impression of Improvement of Foot Pain

An assessment of PGI-I of foot pain will be conducted by the subject using a 7-point Likert scale (from -3: very much worse to +3: very much improved). The PGI-I will be assessed by the subject answering the following question: “Compared to your foot pain prior to the study treatment initiation, your foot pain while performing physical activities (e.g. standing, walking or running) now is: +3=very much improved; +2=much improved; +1=minimally improved; 0=no change; -1=minimally worse; -2=much worse; -3=very much worse”). Also, an additional question will be included in the PGI-I assessment which will ask patients the following: "If you experienced a change, was this change meaningful to you? (Yes/No)"

Patient Global Impression of Severity of Foot Pain

An assessment of PGI-S of foot pain will be conducted by the subject using a 4-point Likert scale (from 0: no pain to 3: severe pain). The PGI-S will be assessed by the subject by answering the following question: “How severe was your foot pain while performing physical activities (e.g. standing, walking or running) over the past week?” (0=no pain; 1=mild pain; 2=moderate pain; 3=severe pain).

Patient Global Impression of Improvement of Disability

An assessment of PGI-I of the subject's disability will be conducted by the subject using a 7-point Likert scale (from -3: very much worse to +3: very much improved). The PGI-I will be assessed by the subject answering the following question: “Compared to your disability prior to the study treatment initiation, your disability while performing physical activities (e.g. standing, walking or running) now is: +3=very much improved; +2=much improved; +1=minimally improved; 0=no change from baseline; -1=minimally worse; -2=much worse; -3=very much worse”). In addition, an additional question will be included in the PGI-I assessment which will ask patients the following: "If you experienced a change, was this change meaningful to you? (Yes/No)".

Patient Global Impression of Severity of Disability

An assessment of PGI-S of disability will be conducted by the subject using a 4-point Likert scale (from 0: no disability to 3: severe disability). The PGI-S will be assessed by the subject by answering the following question: "How severe was your disability while performing physical activities (e.g. standing, walking or running) over the past week?" (0=no disability; 1=mild disability; 2=moderate disability; 3=severe disability).

After protocol amendment 3 (September 07, 2018), the collection period for PGI went from daily 7 days prior to the visit to only being collected at the visit. For subjects who have the diary data collected daily 7 days prior to the visit, the last value on or before (within 7 days prior to the visit) will be used for the result at the visit for summary and analysis. All the diary data will be provided in the listing.

Endpoints and corresponding analyses are described in Table 4.

Table 4 PGI - Endpoints and Corresponding Analyses

Endpoint	Analyses
Mean change from baseline to each post-treatment visit in the PGI-S pain and disability scores at each post-treatment visit	- Frequency table will be presented at all timepoints. DB: - ANCOVA at Week 4, Week 8 and Week 12 (separately). This model will include treatment group, and the stratification parameter as fixed categorical covariates and the baseline value as a continuous covariate. Active Treatment Cycle: - A summary by number of active treatment cycles
Mean PGI-I pain and disability scores at each post-treatment visit	- Frequency table will be presented at all timepoints. DB: - ANCOVA at Week 4, Week 8 and Week 12 (separately). This model will include treatment group, and the stratification parameter as fixed categorical covariates. The additional question will be summarized and listed but will not be used for statistical analysis. Active Treatment Cycle: - A summary by number of active treatment cycles

5.7.3.4 HV angle and intermetatarsal angle

Endpoints and corresponding analyses are described in Table 5.

Table 5 HV Angle and Intermetatarsal Angle - Endpoints and Corresponding Analyses

Endpoint	Analyses
Mean change from baseline to each post-treatment visit in the HV angle as measured directly by weight-bearing anterior-posterior radiographs	- descriptive statistics: actual values and change from baseline values will be presented at all timepoints. DB: - MMRM at Week 4, Week 8 and Week 12. MMRM setup is similar to the primary analysis. Active Treatment Cycle: - A summary by number of active treatment cycles
Mean change from baseline to each post-treatment visit in intermetatarsal angle as measured directly by weight-bearing anterior-posterior radiographs	

5.7.3.5 Quality of life (SF-36)

The SF-36 is a 36-item questionnaire which measures the extent to which physical health impacts an individual's functional ability and perceived well-being in mental, social and physical aspects of life. The SF-36 has eight individual subscales divided across physical and psychological domains: physical function, role physical, bodily pain, global health, vitality, social function, role emotional and mental health. Scores on these subscales can be combined to form two higher-order summary scores, the Physical Component Summary and Mental Component Summary. The Physical Component Summary is calculated by positively weighting the four subscales in the physical domain (physical function, role physical, bodily pain and global health) and the remaining psychological domain subscales negatively. In contrast, the Mental Component Summary is calculated by positively weighting the four mental domain subscales (mental health, vitality, social function and role emotional), and negatively weighting the four physical domain subscales.

Endpoints and corresponding analyses are described in Table 6

Table 6 SF-36 - Endpoints and Corresponding Analyses

Endpoint	Analyses
Mean change from baseline to each post-treatment visit in quality of life as measured by the SF-36 (all 8 domains and components scores: Physical Component Summary and Mental Component Summary, with focus on Physical Component Summary, physical function and bodily pain)	The 8 domain subscale scores and 2 component scores of the SF-36 will be derived by OPTUM - descriptive statistics: actual values and change from baseline values of each domain and components score will be presented at all timepoints. DB: - ANCOVA at Week 8 and Week 12 (separately). This model will include treatment group and the stratification parameter as fixed categorical covariates and the baseline value as a fixed continuous covariate. Active treatment cycle: - A summary by number of active treatment cycles will be used.

5.7.3.6 Time to retreatment

Time to retreatment will be calculated as ((Date of retreatment – previous injection date) + 1) / 7 (week). Subjects who are not retreated are considered censored at the date of their last visit (follow-up). Log rank test adjusting for the stratification will be performed to compare the mean time to retreatment between placebo group and each Dysport group in the DB period. This analysis will be performed on the safety population.

Time to retreatment will be calculated for each cycle. Descriptive statistics will be provided for each active cycle.

5.7.4 Analysis of Exploratory Efficacy Endpoints

CCI

CCI [redacted]

5.7.5 Analysis of Additional Endpoints

The supplementary analysis (1) in section 5.7.2.4 will be repeated for Week 4 and Week 12 DB:
Proportion of responders in the NPRS pain score at Week 4 and Week 12 DB. A responder is defined as a subject with 20% decrease from baseline (30%, 40% and 50% cut-off will also be considered).

5.8 Safety

5.8.1 General Consideration

All safety data will be included in subject data listings (see listing detail conventions in Appendix A5). There will be no statistical comparison between the treatment groups for safety data, unless otherwise specified within the relevant section.

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, some summary tables will be also provided by dose received at each active treatment cycle on the Active Treatment population, this will be specified in the dedicated section.

5.8.2 Extent of exposure

The duration of exposure will be calculated:

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

The following extent of exposure summaries will be presented:

- Summary of duration of study treatment exposure will be provided by double-blind treatment group and overall subjects from the safety population.
- Summary of duration of treatment cycle, for each treatment cycle, dose received in each treatment cycle and overall subjects from the safety population.
- A summary table with the total number of subjects who received 1, 2 or 3 active treatment cycles and overall subjects, all doses combined on the active treatment population.

- A summary table with the total number of subjects in the Safety population by sequence of dose will be provided.

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

- Description of the actual dose in DB period by DB treatment group from the safety population.
- Actual dose will be described for each dose levels.
- Description of the actual dose at each treatment cycle, by treatment group from the safety population.
- Actual dose will be described for each dose levels.
- A listing of exposure data will also be provided.

5.8.3 Adverse Event

All adverse events (AEs) recorded in the eCRF will be coded using the latest version of MedDRA dictionary in effect within IPSEN at the time of the database lock. AEs will be classified as treatment-emergent AEs (TEAEs) according to the rules below:

- Events with start date on or after the date of first IMP administration.
- Events whose severity worsens on or after the date of first IMP administration,
- Refer to appendix A3 for handling of partial date. In the case where it is not possible to define an AE as treatment emergent or not, the AE will be classified by the worst case; i.e. treatment emergent.

The following AE summaries will be provided:

- An overview table summarizing the number and percentage of subjects with at least one of the following AEs: any AE; any TEAE; treatment-related TEAE; TEAE leading to discontinuation from the study, SAE, treatment-related SAE, SAE leading to death, AESI, (this summary table will also be provided by dose received at each active treatment cycle),
- A summary of the number and percentage of subjects reporting a TEAE by treatment group, SOC and PT, (this summary table will be also be provided by dose received at each active treatment cycle),
- A summary of the number and percentage of subjects reporting a TEAE by treatment group and PT,
- A summary of the number and percentage of subjects reporting a TEAE by severity, SOC and PT,
- A summary of the number and percentage of subjects reporting a TEAE by treatment group, causality, SOC and PT, and
- A summary of non-serious TEAE by treatment group, SOC and PT.

For DB period, AEs summaries will be ordered in term of decreasing frequency for SOC and PT within SOC in the Dysport 500 U group and then similarly by decreasing frequency in the Dysport 300 U group, and then alphabetically for SOC and PT within SOC.

AEs will be counted as follows:

- Subjects with more than one AE within a particular SOC are counted only once for that SOC. Similarly, subjects with more than one AE within a particular PT are counted only once for that PT;

- Subjects reporting a TEAE more than once within that SOC/ PT, the TEAE with the worst case severity will be used in the corresponding severity summaries;
- Subjects reporting a TEAE more than once within that SOC/ PT, the TEAE with the worst case relationship to study medication will be used in the corresponding relationship summaries;
- If the severity is missing for a TEAE, it will be considered as missing in the summary tables;
- Summary by intensity will be presented on that order: severe > missing > moderate > mild;
- If the causality is missing for a TEAE, it will be considered related in the summary tables.

A listing with all AE data will be listed by treatment group including non-TEAEs. Treatment-emergence status will be flagged.

Deaths, Serious Adverse Events, and Other Significant Adverse Events

The following summary tables will be provided if more than 5 subjects satisfy the corresponding criteria:

- A summary of the number and percentage of deaths during the study, by treatment group, SOC and PT,
- A summary of the number and percentage of subjects reporting a serious TEAE, by treatment group, SOC and PT,
- A summary of the number and percentage of subjects with AEs leading to discontinuation of study treatment, by treatment group, SOC and PT.

Adverse events of special interest (AESIs) for Dysport are TEAEs that suggest a possible remote spread of effect of the toxin 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 PTs compatible with the mechanism of action of BTX-A-HAC 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. A list of MedDRA preferred terms, used to identify any potential AESI, is provided in a separate document 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 'hypersensitivity reactions' due to study treatment administration. Cases will be excluded if they are confounded by presence of alternative clinical etiologies (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 summaries, only the final list of AESIs confirmed by the sponsor as "a possible remote spread event" or "hypersensitivity reactions" will be taken into account.
The following AESI summary will be provided.

- AESIs with the number and percentage of subjects and the number of occurrences presented by primary SOC and PT, presented by type of AESI.

The following listings will also be provided.
- A listing of all deaths that occurred during the study,
- A listing of all serious adverse events,
- A listing of all adverse events leading to study discontinuation,
- A listing of subjects having AESIs.

5.8.4 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 A2,

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 patients with abnormal result of drug abuse test.

5.8.5 Vital Signs

The following summaries will be provided.
- A summary of the actual and change from baseline in each vital sign parameter by treatment group,
- A summary of the number and percentage of subjects experiencing PCSA. PCSA criteria are defined in appendix A2.

The following listing are to be provided:
- A listing of vital sign data by treatment group, with abnormal value highlighted and unscheduled vital signs flagged,
- Listing of PCSA. All data for a vital sign parameter will be displayed for a subject having at least one post-baseline PCSA (with flag indicating PCSA).

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

5.8.6 Physical Examination

The following summary and listings will be provided:

- A shift from baseline (normal vs abnormal) to each post-baseline visit
- A listing of physical examination data, with unscheduled visit flagged
- A listing with any subjects with at least one physical examination abnormality.

5.8.7 Examination of Injected Foot

The following summary and listings will be provided:

- A summary of each parameter for the dermatologic exam, neurological exam, and musculoskeletal exam at each visit by treatment group,
- A separate listing will be provided for all subjects and results, with unscheduled visit flagged [U].
- A listing of examination data, with unscheduled visit flagged [U], for subjects with at least one abnormal result. All results will be displayed for the subject if one result is abnormal.

5.9 Pharmacokinetics and Pharmacodynamics

Pharmacokinetics and pharmacodynamics are not assessed in this study.

5.10 Anti-drug 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.

The following conventions will be considered:

- 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 positive baseline results,
  - quartiles of the titer range,
  - number of seroreverters,
  - number of treatment-boosted,
  - number and percentage of seroconverters,
  - incidence.

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 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 titer > 1 post baseline. This listing will include the combined information about baseline BTX status, primary efficacy results and TEAEs,

6 DATA HANDLING

6.1 Visit Window

No time window for visit/timepoints will be defined.

6.2 Unscheduled Visits, Retest, Withdrawal Visit

All listings will include retests and unscheduled visits, while for the description by visit in the tables, only the scheduled visits according to the protocol will be described. Unscheduled visit and retest measurements will be used to provide a measurement for a baseline data or endpoint value (e.g. worst value), if appropriate according to their definition. These measurements will also be used to determine abnormal laboratory or vital signs values.
If a value requires a retest (for laboratory values or vital signs) the closest non-missing reliable value to the scheduled visit will be used in the summary tables.
Subjects who have withdrawn early from the study have their last assessment entered in the Case Report Form (CRF). By convention, for these subjects, the visit number will be reassigned to the next empty visit number.

7 DERIVED DATA

All the derived data are defined in the corresponding sections.

8 REFERENCES

Reference to ICH regulatory guidelines:

- ICH E3: Structure and Content of Clinical Study Reports
- ICH E6: Good Clinical Practice
- ICH E9: Statistical Principles for Clinical Trials
- ICH E9 (R1) Addendum: Estimands and Sensitivity Analysis in Clinical Trials

Reference to EMA or point to consider guidelines:

- Adjustment for baseline covariates in clinical trials
- Choice of a non-inferiority margin
- Clinical trials in small populations
- Data monitoring committees
- Investigation of subgroups in confirmatory clinical trials

- Missing data in confirmatory clinical trials
- Application with Meta Analyses, One pivotal study
- Multiplicity issues in clinical trials
- Switching between superiority and non-inferiority

Reference to FDA guidelines:

- Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics

Standard SDTM+ user guide

References for ADA:

- ABIRISK Terms and Definitions for Reporting Immunogenicity Results

9 APPENDICES

A1. SAS code

SAS code for primary efficacy analysis:

```
*********************************************************************
** CHG  = Change from Baseline to the specific visit in foot pain**
** Trt  = Treatment group, Placebo, Dysport 300U and 500U        **
** Visit= Week 4, Week 8 and Week 12                             **
** Base = Baseline pain                                         **
** Type = Stratification parameter, unilateral and bilateral HV  **
** Usubjid = Subject id                                         **
*********************************************************************
PROC MIXED;
   CLASS Trt(ref="Placebo") Type Visit(ref="Week8") Usubjid;
   MODEL CHG = Base Trt Type Visit Trt*Visit / CL ALPHA=0.05;
   LSMEANS Trt / PDIFF CL;
   REPEATED Visit / TYPE=UN SUB=Usubjid;
RUN;
```

SAS code for Supplementary analysis (1):

```
*********************************************************************
** Responder = 1 if is Responder, otherwise zero                **
** Trt  = Treatment group: Placebo, Dysport 300U and 500U        **
** Base = Baseline pain                                         **
** Type = Stratification parameter, unilateral and bilateral HV  **
*********************************************************************
PROC LOGISTIC;
   CLASS Trt Type;
   MODEL Responder = Base Type Trt / EXPB;
   ODDSRATIO Trt;
   LSMEANS Trt / e diff oddsratio cl;
RUN;
```

A2. List of PCSA criteria

PCSA for Vital Signs parameters:

Parameter	PCSA
Systolic Blood Pressure	≤ 90 mmHg and change from baseline ≤ -20 mmHg ≥ 180 mmHg and change from baseline ≥ 20 mmHg
Diastolic Blood Pressure	≤ 50 mmHg and change from baseline ≤ -15 mmHg ≥ 105 mmHg and change from baseline ≥ 15 mmHg
Heart Rate	≤ 50 bpm and change from baseline ≤ -15 bpm ≥ 120 bpm and change from baseline ≥ 15 bpm
Weight	≤ -5% change from baseline ≥ 5 % change from baseline

PCSA for blood chemistry parameters:

Parameter	PCSA
ALT/SGPT	≥ 3 x ULN IU/L
AST/SGOT	≥ 3 x ULN IU/L
GGT	≥ 3 x ULN IU/L
ALP (Alkaline Phosphatase)	≥ 1.5 x ULN IU/L or ≥ 3 x ULN IU/L
Creatinine	≥ 150 µmol/L (Adults)
Blood Urea Nitrogen	> 10 mmol
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
Lipase	≥ 3 x ULN U/L
Amylase	≥ 3 x ULN U/L
Glucose	<2.7 mmol/L > 11 mmol/L (unfasted)
HbA1c	> 8%
Albumin	≤ 25 g/L
Direct bilirubin	> 2 ULN µmol/L
Total bilirubin	> 2 ULN mmol/L
Phosphorous	< 0.75 mmol/L > 1.8 mmol/L
Total protein	< 40 g/L

PCSA for haematology parameters (proposal - PCSA criteria need medical input)**:**

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
Neutrophilis	< 1 10^9/L > 13 10^9/L
Monocytes	> 2.3 10^9/L
Basophilis	10^9/L
Eosinophilis	10^9/L
Hemoglobin	≤ 95 g/l (female) ≤ 115 l/l (male)
Platelet count	<75.1 10^9/L ≥ 700 10^9/L

PCSA for coagulation parameters:

Parameter	**PCSA**
INR	>3
Prothrombin Time (PT)	> 17.1 seconds
Partial Thromboplastin Time (PTT)	> 59 seconds

A3. Partial/Missing Date Convention

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 / partial, it is assumed to have occurred during the study treatment phase (i.e. a TEAE for AEs) except if the partial onset date or the stop date indicates differently.
- 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 first IMP 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).

Algorithm for Prior/Concomitant

Medication, non-drug 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 only, Prior and Concomitant, or Concomitant only.

In case of partial start and/or stop medication/ non-drug therapies/surgical procedures dates, imputation will be done to determine the classification:

- If a partial start date, the first day of the month will be imputed for missed day and January for missing month,
- If a partial stop date, the last day of the month will be imputed for missed days and December will be imputed for missing month.

In case incomplete start or stop date does not allow the classification, the medication, non-drug therapy, or surgical procedure will be classified as concomitant.

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).

Algorithm for TEAE

For deriving the TEAE flag the following process of temporary date imputation is done (for AE start date only assuming no AE end date are missing). The date imputation algorithm for incomplete adverse event start dates is described in Table A1. Classification of adverse event according to its treatment-emergent status is then done using the imputed date.

In the following table, all dates are presented using an YYYY-MM-DD format. As an example, suppose First IMP administration = 2002-08-11 and several AEs have incomplete start dates.

Table A1: Data imputation algorithm for AE start date (AESTDT)

Description of incomplete date	Imputed numeric date	Example	
		Character date	Imputed date
Day is missing			
YYYY-MM < YYYY-MM of [First IMP admin.]	YYYY-MM-01	2002-07-XX	2002-07-01
YYYY-MM = YYYY-MM of [First IMP admin.]	Min ([First IMP admin.], AE end date)	2002-08-XX	Min (2002-08-11, AE end date)
YYYY-MM > YYYY-MM of [First IMP admin.]	YYYY-MM-01	2002-09-XX	2002-09-01
Day and month are missing			
YYYY < YYYY OF [First IMP admin.]	YYYY-01-01	2001-XX-XX	2001-01-01
YYYY = YYYY OF [First IMP admin.]	Min ([First IMP admin.], AE end date)	2002-XX-XX	Min (2002-08-11, AE end date)
YYYY > YYYY OF [First IMP admin.]	YYYY-01-01	2003-XX-XX	2003-01-01
Day, month, and year are missing			
XXXX-XX-XX	Min ([First IMP admin.], AE end date)		Min (2002-08-11, AE end date)

YYYY = non-missing year, MM = non-missing month, DD = non-missing day, XX = missing field.

If an AE onset date is partial or missing, the event will be allocated to the first IMP administration where onset could have occurred (taking into account date and time stopped).

A4.Programming Convention for Outputs

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.

The mean, median, lower quartile, upper quartile, SD and standard errors of the mean (SE), 95% confidence interval values will be reported to one decimal place greater than the raw data recorded in the database.

The minimum and maximum values will be reported with the same number of decimal places as the raw data recorded in the database.

In general, the maximum number of decimal places reported should be four for any summary statistic.

Percentages will be presented to one decimal place. Percentages will not be presented for zero counts. Percentage will be calculated using n as denominator. The denominator will be specified

in a footnote for clarification if necessary. If sample sizes are small (less than 15), the data displays will show the percentages, but in the CSR only frequencies should be described.

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. In summary tables or analysis, data like "<a" will be replaced with "0.5*a", data like ">b" will be replaced with "b".

Dates will be presented in the format [ddmmmyyyy] and times in the format [hh:mm].

A5. Listings conventions

Any listings will contain at least the following data: subject identifier, age and gender. When dates are presented, the associated study days should be included. They should be sorted by treatment group then subject identifier. Listings should be broken down by centre and treatment group.