CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

125360

STATISTICAL REVIEW(S)

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Biostatistics

Statistical Review and Evaluation
CLINICAL STUDIES

NDA/Serial Number:	BLA125360
Drug Name:	NT 201 (Xeomin)
Indication(s):	Treatment of Cervical Dystonia
Applicant:	MRZ
Date of Document:	July 1, 2009
Biometrics Division:	Division 1 (HFD-710)
Statistical Reviewer:	Ohidul Siddiqui, Ph.D
Concurring Reviewers:	Kun Jin, Ph.D; Kooros Mahjoob , Ph.D
Medical Division:	HFD-120
Medical Reviewer:	Anne Constantino
Project Manager:	Vandna N Kishore

Keywords: *Clinical studies, NDA review, statistical review, endpoint analysi/, multi-center, wilcoxon test*

Table of Contents

1. Executive Summary

The sponsor has claimed an efficacy evidence of NT 201 in the treatments of cervical dystonia (CD) and blepharospasm (BEB) based on two randomized, double-blind, placebo-controlled, Phase III trials (Studies 0408/1, and 0433/1). As a supportive efficacy evidence of NT 201, the sponsor has submitted efficacy findings of another two placebo control randomized studies (Studies 0013 and 0003/1). Tables 1 and 2 list the design features and the study population features of the studies.

1.1. Conclusions and Recommendations

Xeomin has demonstrated its efficacy in treating patients with cervical dystonia. The comparison of each Xeomin group to the placebo group was statistically significant at $p<0.001$. Subgroup analyses also revealed that the efficacy of the two Xeomin doses was similar in pre-treated subjects (i.e., subjects who had received a Botulinum toxin prior to this study) and in naïve subjects (i.e., subjects who had not received a Botulinum toxin prior to this study).

Xeomin also has demonstrated its efficacy in treating patients with benign essential blepharospasm. Comparison of the Xeomin group to the placebo group was statistically significant at $p<0.001$.

1.2. Brief Overview of Reviewed Clinical Studies

A Phase 3 (Studiy#408/1), randomized, double-blind, placebo-controlled, multi-center trial in a total of 233 subjects with cervical dystonia was conducted to investigate the effectiveness of Xeomin. The subjects who had a clinical diagnosis of predominantly rotational cervical dystonia (spasmodic torticollis) were randomized in the study. Subjects were randomized (1:1:1) to receive a single administration of Xeomin 240 Units (n=81), Xeomin 120 Units (n=78), or placebo (n=74). Each subject received a single administration of 4.8 mL of reconstituted study agent (Xeomin 240 U, Xeomin 120 U, or placebo). The primary efficacy endpoint was the change in the Toronto Western Spasmodic Torticollis Rating Scale (TWSTRS) total score from baseline to Week 4 post injection. The primary analysis of the primary efficacy measure was based on the comparison of least square (LS) means from an analysis of covariance (ANCOVA) model at Week 4 between treatment groups. The ANCOVA model includes treatment, Baseline TWSTRS-Total score, gender, age, pre-treatment of CD with a Botulinum toxin, and pooled center. Missing data for the change from Baseline of the TWSTRS-Total score were replaced with the subject's baseline value (no change).

For the multiplicity adjustment, pairwise comparisons between treatment groups were performed by using a fixed-sequence test procedure (step downward) in the ITT Population

starting with the comparison of the 240 U group vs. placebo followed by the comparison of the 120 U group vs. placebo.

Another Phase 3 (Study#0433/1), randomized, double-blind, placebo-controlled, multi-center trial was conducted to investigate the efficacy of Xeomin in treating in subjects with benign essential blepharospasm. Subjects who had a clinical diagnosis of bilateral benign essential blepharospasm, with baseline Jankovic Rating Scale (JRS) Severity subscore ≥ 2 were randomized in the trial. Subjects were randomized (2:1) to receive a single administration of Xeomin (n=75) or placebo (n=34). The primary efficacy variable was the change from Baseline to Week 6 after injection in the JRS Severity subscore (rated by an independent investigator blinded to the subject's treatment assignment). The primary analysis of the primary efficacy measure was based on the comparison of LS means from an ANCOVA model at Week 6 between the two treatment groups in the ITT Population. The dependent variable in the ANCOVA model was the change from Baseline in the JRS Severity subscore and the independent variables were treatment, Baseline JRS Severity subscore, gender, age, dose group, and pooled center. The last observation carried forward (LOCF) approach was used for dealing with missing data.

As a supportive efficacy evidence of NT 201, the sponsor also submitted efficacy findings of another two placebo control randomized study (Studies 0013 and 0003/1).

1.3. Statistical Issues and Findings

No statistical issues were found in the reviewed studies.

2. Introduction

The sponsor has claimed an efficacy evidence of NT 201 in the treatment of cervical dystonia (CD) and blepharospasm (BEB) based on two randomized, double-blind, placebo-controlled, Phase 3 trials. As a supportive efficacy evidence of NT 201, the sponsor has submitted efficacy findings of another two placebo control randomized study (Studies 0013 and 0003/1). Tables 1 and 2 list the design features and the study population features of the studies.

Table 1. Designs of the Pivotal and Supportive Studies

Study No. #	Study Objective(s)	Design and Type of Control	Test Product(s); Dosage Regimen; Duration	Patients in ITT Population (completed) Patients per treatment group	Major Endpoints
Cervical Dystonia					
0408/1 37 centers USA completed (Jul 06/Mar 08)	Safety and efficacy of two NT 201 doses compared with placebo in pre-treated and treatment-naive patients with CD	Phase 3, prospective, double-blind, randomized, placebo-controlled, multicenter.	120 U or 240 U NT 201, or placebo, in 1:1:1 ratio One IM injection with follow-up for 8 to 20 weeks.	Pre-treated and treatment-naive1 patients with CD ITT: n=233 (219) NT 201 240 U: n=81 NT 201 120 U: n=78 placebo: n=74	Primary: Change from Baseline in TWSTRS-Total score at Week 4. Secondary: Change from Baseline to post-Baseline visit in TWSTRS-Disability, TWSTRS-Severity and TWSTRS-Pain scores
0013 /1 51 centers Belgium, Czech Republic, France, Russia, Germany, Slovakia, Sweden, Austria, Poland, Hungary, Israel	Safety and efficacy of NT 201 compared with Botox in pre-treated patients with CD	Phase 3, prospective, double-blind, randomized, non-inferiority, active-controlled, multicenter study	70-300 U NT 201 vs. Botox One IM injection with follow-up for up to 16 weeks	Patients with CD successfully pre-treated1 on stable dose of Botox ITT: 463 (451) NT 201: 231 Botox: 232	Primary: Change from Baseline in mean TWSTRS-Severity Score at Day 28. Secondary: Change from Baseline in TWSTRS-Pain, TWSTRS-Factorial, VAS Pain, and PEGR at Day 28 and Final Visit; Change from Baseline in TWSTRS-Severity score at Final Visit; Response analysis at Day 28; Investigator's global assessment of efficacy at Final Visit

Benign Essential Blepharospasm (BEB)					
0433/1 19 centers USA, Canada	Safety and efficacy of NT 201 compared with placebo in pretreated patients with BEB	Phase 3, prospecti ve, double blind, randomi zed, placebo-controlle d, multicen ter study	Up to 50 U NT 201 per eye vs. placebo Main Period: One IM injection with follow-up for up to 20 weeks	Pre-treated1 patients with BEB; ITT: 109 (102) NT 201: 75 placebo: 34	Primary: Change from Baseline to Week 6 in JRS Severity subscore (assessed by independent rater) Secondary: Change from Baseline to Week 6 in JRS Severity subscore (assessed by patient); Blepharospasm Disability Index at Week 6
0003 42 centers Belgium, Czech Republic, France, Germany, Hungary, Israel, Poland, Russia, Slovakia	Safety and efficacy of NT 201 compared to Botox in subjects with BEB	Phase 3, double blind, randomi zed, non-inferiorit y, active-controlle d, multicen ter study	Up to 70 U NT 201 vs. Botox One IM injection with follow-up for up to 16 weeks	Subjects with BEB, successfully pre-treated1 with Botox on a stable dose; ITT: 300 (294) NT 201: 148 Botox: 152	Primary: Change from Baseline in JRS sumscore at Day 21 Secondary: Change from Baseline to Final Visit in JRS sumscore; Investigator's global assessment of efficacy at Final Visit; time to onset and waning of effect; duration of effect

Source: Summary of clinical efficacy report

Table 2: Key Features of the Study Populations in the Phase 3, Placebo-Controlled Studies of NT 201

	Study 0408/1	**Study 0433/1**	**Study 0013 /1**	**Study 0003/1**
Pre-Treated with BoT	61%	100%	96%	100%
Gender	34% male	35% male	38.2% male	27.3%male
Mean age (±SD)	53 (11.5)	62 (10.3)	49.7 (11.9)	62.7 (10.23)
Race	91% Caucasian	83% Caucasian	100% Caucasian	100% Caucasian

Source: Summary of clinical efficacy report

Disposition of Patients

Majority of patients completed the studies as planned (Table 3). The main reasons of discontinuation from the studies were Withdrawal criteria occurred and treatment-unrelated adverse events.

Table 3. Patient Disposition

Study	Treated Patients (N)	Completed N (%)	Discontinued Due to Lack of Efficacy: N (%)
0408/1	233	219 (94.0)	14 (6.0)
0433/1	109	102 (93.6)	7 (6.4)
0013/1	463	451 (97.4)	5 (1.1)
0003/1	300	294 (98%)	--

Source: study reports

Data Sources

The study reports and SAS data sets are available at
\\cbsap58\M\eCTD_submissions\STN125360\0000\m5\datasets

3. Statistical Evaluation

Demographic and Baseline Characteristics

The demographic characteristics were similar across treatment groups with no statistically significant differences within each study. The mean ages across the studies were in the range of 53-62 years. Distribution of female patients across the studies was in the range of 62% to 66%. The majority of patients were Caucasians.

Efficacy Evaluation

Study 408/1 (CD)

Study 408/1 was a Phase 3, prospective, double-blind, randomized, placebo-controlled, multicenter study. The randomized patients were adults 18 to 75 years of age with CD of the predominantly rotational form (i.e., spasmodic torticollis). Both pre-treated and treatment-naive patients were eligible for the study; the study protocol called for at least 40% of the study patients to be treatment-naive. Pre-treated patients must have had a stable response to the two most recent injections, with a maximum dose of 300 U Botulinum toxin type A or 12,000 U Botulinum toxin type B per injection, and must not have received an injection within 10 weeks of study Baseline. All patients were required to have the following Toronto Western Spasmodic Torticollis Rating Scale (TWSTRS) scores: Total $\geq$20, Severity $\geq$10, Disability $\geq$3 and Pain $\geq$1.

Treatments:

Patients were randomized (1:1:1) to receive a single intramuscular (IM) administration of 120 or 240 U of NT 201, or placebo. The number and sites of the injections were to be determined by the Investigator. Each patient received a single IM dose of blinded study medication on Day 0. A telephone contact was made on Day 7 following injection, and control visits took place 4 weeks and 8 weeks following injection. In the Main Period of the study, patients were followed for 8 to 20 weeks, until a new injection was required.

Primary Efficacy Measure

The primary efficacy variable was the change from Baseline to Week 4 following injection in the TWSTRS-Total Score, in the Intent-to-Treat (ITT) Population. The primary analysis of the primary efficacy measure was based on the comparison of least square (LS) means from an analysis of covariance (ANCOVA) model at Week 4 between treatment groups. The ANCOVA model includes treatment, Baseline TWSTRS-Total score, gender, age, pre-

treatment of CD with a Botulinum toxin, and pooled center. Missing data for the change from Baseline of the TWSTRS-Total score were replaced with the subject's baseline value (no change).

For the multiplicity adjustment, pairwise comparisons between treatment groups were performed by using a fixed-sequence test procedure (step downward) in the ITT Population starting with the comparison of the 240 U group vs. placebo followed by the comparison of the 120 U group vs. placebo.

For the sensitivity analyses, two other models were presented: the final model with all adjusting variables with an influence of $p \leq 0.2$ on the model (backward selection), and the simple model including only the treatment effect.

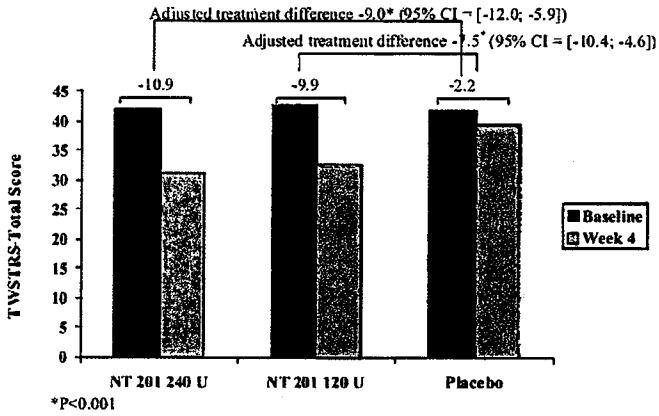
Secondary Efficacy Variables:

The secondary efficacy variables were changes from Baseline to all post-Baseline visits in: TWSTRS-Total score- Pre-treated and Treatment-naive Subjects, TWSTRS-Disability, TWSTRS-Severity and TWSTRS-Pain scores, and Global Assessment by Investigator.

Efficacy Findings:

According to the primary efficacy analysis in the ITT population (using patient's baseline value as a replacement for missing value), the change in TWSTRS-Total score from Baseline to Week 4 was significantly greater in the NT 201 groups, compared with the placebo group ($p<0.001$), irrespective of the statistical model or replacement strategy used. The least square (LS) mean difference between the change in each NT 201 group and placebo was highly statistically significant ($p<0.001$; ANCOVA) and clinically meaningful: -9.0 points for 240 U vs. placebo, and -7.5 points for 120 U vs. placebo (Fig. 1). The difference between each NT 201 dose group and the placebo group persisted at Week 8 and the Final Visit of the Main Period. Since the dropout rate is minimal, both doses were statistically significantly ($p<0.0001$) efficacious as compared to placebo at week 4 in the observed cases analysis. Using different approaches in dealing with missing data (e.g., OC, LOCF, and MMRM analyses, etc.) have no impact on the significance of the two doses compared to placebo at the end of week 4. For the primary analyses, comparisons between the treated groups were performed by using a fixed-sequence test procedure (step downward). 1st step: 240 U versus placebo. 2nd step: 120 u versus placebo.

Figure 1: Mean TWSTRS-Total Score at Baseline and Week 4 and Respective Score Differences by Treatment Group (Full Model; ITT Population; Study 0408/1 [CD]



Note: Adjusted treatment differences are based on least square (LS) means. Missing values replaced by patient's Baseline value. Source: Study report

Secondary Efficacy Variables

TWSTRS-Subscale Scores

Table 4 lists the ANCOVA results for changes from baseline to Week 4 in the TWSTRS Subscale Scores. For all TWSTRS subscales, both NT 201 doses were superior to Placebo. Reductions in mean scores from baseline to Week 4 were significantly greater in the 240 U and 120 U groups than in the Placebo group ($p<0.003$). At Week 8 and Final Visit, both NT 201 doses were also superior to Placebo in the TWSTRS Subscale Scores.

Table 4: Mean Changes from Baseline to Week 4 in TWSTRS-Subscale Scores (ITT Population; Missing Values Replaced by Baseline Value)

Comparison	TWSTRS Severity score	TWSTRS Disability score	TWSTRS Pain score
	LS mean treatment difference (p-value)	LS mean treatment difference (p-value)	LS mean treatment difference (p-value)
240 U vs. Placebo	-3.9 (<0.001)	-2.8 (<0.001)	-2.2 (<0.001)
120 U vs. Placebo	-2.1 (0.003)	-2.9 (<0.001)	-2.2 (<0.001)

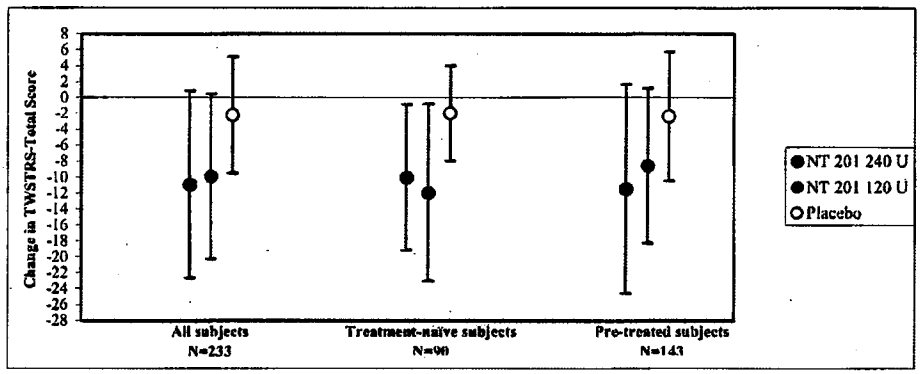
Source: Study report; p values are for "Treatment" and (full model) LS= Least Square,

Subgroup analyses in pretreated and treatment-naive patients

TWSTRS-Total Score- Pre-treated and Treatment-naive Subjects

The mean TWSTRS-Total Scores in pre-treated and treatment-naive subjects (ITT sample) were similar across treatment groups. In pre-treated subjects, the mean TWSTRS-Total score change from Baseline to Week 4 was -11.4 points in the 240 U NT 201 group, compared with -8.5 points in the 120 U group. In treatment-naive subjects, the mean changes were -10.0 points in the 240 U group, and -11.9 in the 120 U group. The clinically relevant change in TWSTRS-Total Score was significantly greater in each subgroup compared with placebo ($p<0.001$ for each comparison). The mean changes all, pre-treated and treatment-naive subjects are presented in Figure 2. Both groups were also significantly (p-value <0.001) different from placebo at week 8 and the final visit.

Figure 2: Mean Change (±SD) in TWSTRS-Total Score from Baseline to Week 4 in All, Pre-treated and Treatment-naive Subjects in Study 0408/1 (CD)



Source: Study report

Table 5 list the mean changes from Baseline to Week 4 in TWSTRS Subscores for Treatment-naive and Pre-treated Patients. In both pre-treated and treatment-naive patients, significant differences in TWSTRS-Severity score change at Week 4 were observed between the 240 U group and the Placebo group. For TWSTRS-Disability score, significant differences in score change were observed in both pre-treated and treatment-naïve patients between the 240 U group and the Placebo group and between the 120 U group and the Placebo group. For TWSTRS-Pain score, significant change differences were observed in pre-treated patients between the 240 U group and Placebo group and in both pretreated and treatment-naïve patients between the 120 U group and the Placebo group. The above findings indicate that there was no dose dependence of the changes in TWSTRS subscores at Week 4.

Table 5: Mean Changes from Baseline to Week 4 in TWSTRS Subscores, Summarized for Treatment-naïve and Pre-treated Patients (ITT Population; Missing Values Replaced by Baseline Value)

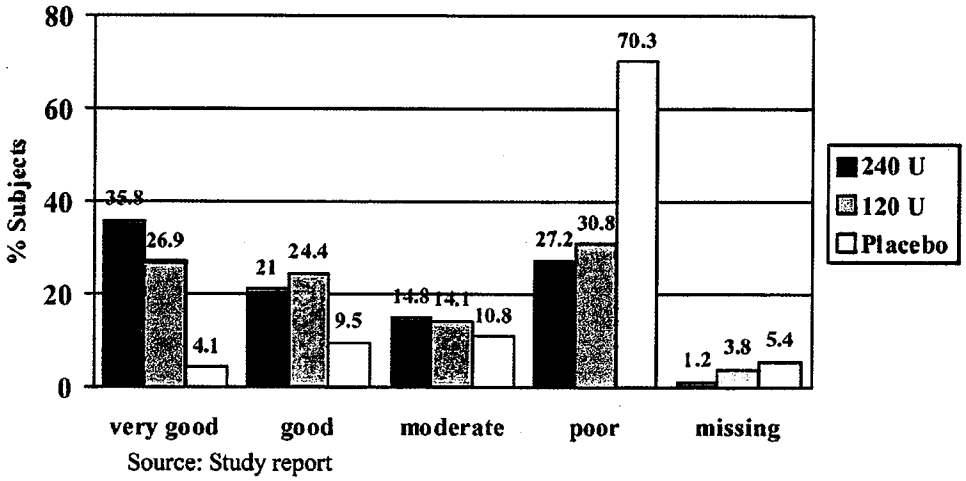
Comparison	Group	LS mean treatment difference	p-value "Treatment" (Full model)
TWSTRS-Severity score			
240 U vs. Placebo	Treatment-naïve Patients	-3.4	0.011
	Pre-treated Patients	-4.4	<0.001
120 U vs. Placebo	Treatment-naïve Patients	-2.0	0.075
	Pre-treated Patients	-1.7	0.051
TWSTRS-Disability score			
240 U vs. Placebo	Treatment-naïve Patients	-3.1	0.001
	Pre-treated Patients	-2.7	0.001
120 U vs. Placebo	Treatment-naïve Patients	-4.7	<0.001
	Pre-treated Patients	-1.9	0.015
TWSTRS-Pain score			
240 U vs. Placebo	Treatment-naïve Patients	-1.3	0.185
	Pre-treated Patients	-3.0	<0.001
120 U vs. Placebo	Treatment-naïve Patients	-2.7	0.006
	Pre-treated Patients	-1.8	0.035

Source: study report; LS= Least Square

Global Assessment by Investigator

Figure 3 lists the global assessment efficacy at the final visit. In the ITT population, the investigator classified the therapeutic efficacy of the 240 U dose and the 120 U dose of NT 201 as very good or good. In the Placebo group "poor" was the most frequent rating.

Figure 3: Global Assessment of Efficacy by Investigator at Final Visit (Total ITT Population)



Source: Study report

Both NT 201 doses (i.e., 120 U and 240 U) were statistically significantly better than placebo based on analyses of the primary endpoint (the reduction of TWSTRS-Total Score) and pre-specified secondary endpoints. Subgroup analyses of the primary efficacy variable indicate the 120 U and 240 U doses of NT 201 were effective in the treatment of CD in both pre-treated and treatment-naive subjects.

Study 0433 / 1 (BEB)

Study 0433/1 was a Phase 3, prospective, double blind, randomized, placebo-controlled, multicenter study. The Patients were randomized in the study who had a clinical diagnosis of bilateral BEB and a JRS Severity subscore of ≥ 2 at Baseline and had a stable clinical response (defined as a consistent, satisfactory response) to at least two previous treatments with a stable dose of Botox (≤ 50 U per eye). A stable dose was defined as a dose similar to the most recent two Botox treatments.

Treatments:

Patients were randomized (2:1) to receive a single intramuscular (IM) administration of NT 201, or placebo. The number and sites of the injections were to be determined by the Investigator. The subjects were followed for 6 to 20 weeks, until a new injection was required.

Primary Efficacy Measure

The primary efficacy variable was the change from Baseline to Week 6 after injection in the JRS Severity subscore (rated by an independent investigator blinded to the subject's treatment assignment). The primary analysis of the primary efficacy measure was based on the comparison of LS means from an ANCOVA model at Week 6 between the two treatment groups in the ITT Population. The dependent variable in the ANCOVA model was the change from Baseline in the JRS Severity subscore and the independent variables were treatment, Baseline JRS Severity subscore, gender, age, dose group, and pooled center. The last observation carried forward (LOCF) approach was used for dealing with missing data.

For the sensitivity analyses, two other models were presented: the final model with all adjusting variables with an influence of $p \leq 0.2$ on the model (backward selection), and the simple model including only the treatment effect.

Secondary Efficacy Variables:

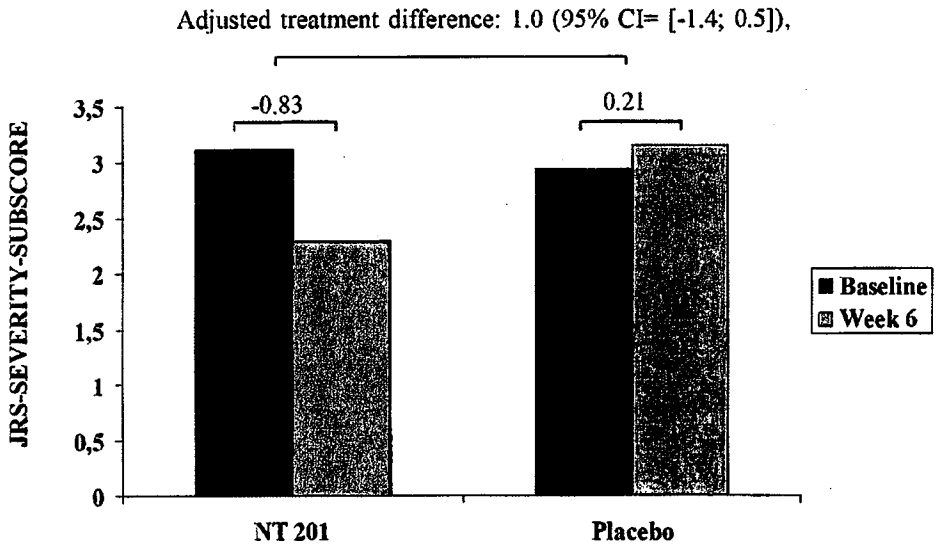
The secondary efficacy variables were change from Baseline at week 6 in the Sumscore (Independent Investigator), change from Baseline in JRS Severity subscore (IVRS), change from Baseline in BSDI , and Patient's Evaluation of Global Response at Final Visit.

Efficacy Findings

The primary efficacy variable was the change from baseline (Day 0) to Week 6 in the JRS Severity subscore as assessed by a blinded Independent Rater. Using the LOCF method for replacement of missing values, the mean change in the JRS Severity subscore from the Baseline to Week 6) in the ITT population was -0.83 points in the NT 201 group and 0.21 points in the Placebo group. JRS Severity subscore values decreased in the NT 201 group from 3.12 points at baseline to 2.29 points at Week 6 and increased in the Placebo group from 2.94 points to 3.15 points (Fig. 4).

For the ITT population, the results of the ANCOVA analyses of the primary variable (based on LS means) showed that the LS mean change from baseline to Week 6 in the JRS Severity subscore was -0.8 points in the NT 201 group and 0.2 points in the Placebo group. The treatment-specific difference of -1.0 points was highly significant ($p<0.001$). Using different approaches in dealing with missing data (e.g., OC, MMRM analyses) have no impact on the significance of the NT 201 compared to placebo at the end of week 6.

Figure 4: Mean JRS Severity Subscore at Baseline and Week 6 and (ITT Population; Missing Values Replaced with LOCF)



*p<0.001

CI=Confidence Interval, JRS=Jankovic Rating Scale

Secondary Efficacy Analyses

The results for the secondary efficacy endpoint analyses are shown in Table 6. For each of the secondary endpoints, the difference between the NT 201 and placebo groups was statistically significant, and the difference was in favor of NT 201.

Table 6: Results for Secondary Efficacy Variables in Study 0433/1 (BEB) Main Period (Full Model; ITT Population)

Variable	Treatment Group	N	Mean	LS Mean Difference (NT 201 – placebo)	p-value [§]
Change from Baseline in JRS Sumscore at Week 6 (Independent Investigator)	NT 201 placebo	75 34	-1.4 0.2	-1.5	<0.001
Change from Baseline in JRS Severity subscore at Week 6 (IVRS)	NT 201 placebo	67 32	-0.8 0.2	-0.8	0.001
Change from Baseline in BSDI at Week 6	NT 201 placebo	75 34	-0.4 0.1	-0.5	0.002
Patient's Evaluation of Global Response at Final Visit	NT 201 placebo	75 34	1.3 -0.6	1.9	<0.001

[§] P-value from full model ANCOVA
Source: Study report

Study 0013 (CD): Supportive study

Study 0013 was a Phase 3, prospective, double-blind, randomized, non-inferiority, active-controlled, multicenter study. The randomized subjects were up to 75 years of age with spasmodic torticollis and TWSTRS scores: Severity ≥10, Severity (rotation) ≥2, and severity score for rotation greater than score for laterocollis, anterocollis or retrocollis. The pre-treated subjects with a stable therapeutic dose of Botox (defined as at least two injections into the same muscles, in the same total doses and volumes, with any time interval between injections differing by ≤3 weeks) were eligible to participate. The most recent Botox injection was required to be at least 10 weeks before randomization.

Treatments:

Subjects were randomized (1:1) for a single IM injection of NT 201 or Botox at the same dose as the most recent dose of Botox (total dose 70 to 300 U, registered in Europe as BOTOX). Subjects were followed for up to 16 weeks.

Primary Efficacy Measure

The primary efficacy variable was the change from Baseline to Week 4 in TWSTRS-Severity score. The primary efficacy variable was analyzed in the treated per-protocol (TPP) population. An analysis of covariance (ANCOVA) was used for the primary efficacy analysis. The dependent variable was the change of the TWSTRS-Severity scale and the independent variables were treatment, baseline TWSTRS-Severity scale, total dose, sex, age, number of injection sessions since diagnosis of torticollis (used a categorical variable), and country. The final model used for statistical inference included all variables and covariates having an influence on the primary efficacy variable with $p < 0.2$. A backward selection method was used for model building purposes. For the primary statistical analysis of the primary efficacy

variable, no missing data were imputed. However, to perform a sensitivity analysis of the influence of missing data on the study outcome in addition to the confirmatory analysis two strategies (replacements by zeros or by group visit means) to handle missing data were applied using the ITT population.

NT 201 was considered as clinically not inferior, if the upper 95% confidence bound mean difference was lower than Δ. Δ was defined to be 1.3 points of the TWSTRS-Severity scale. If the upper 95% confidence bound was less than zero, NT 201 would be declared superior to BOTOX .

Secondary Efficacy Variables:

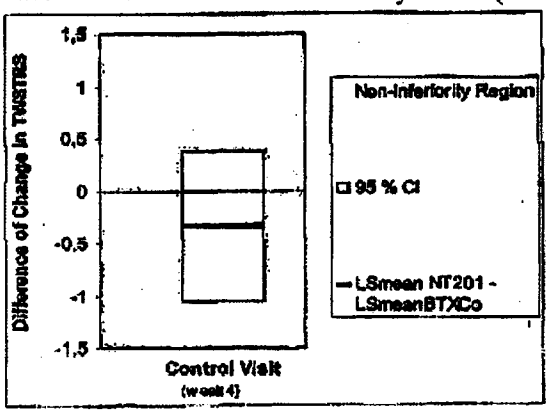
The secondary efficacy variables were the TWSTRS-Pain subscale, the VAS pain scale, Investigator's Global Assessment of Efficacy. Two-sample Wilcoxon test was used to analyze the secondary endpoints.

Efficacy Findings

The patient subset used for the confirmatory efficacy analysis in this study was defined as the "treated per-protocol (TPP)" sample of randomized study patients. The LS mean difference in the changes of TWSTRS-Severity score from Baseline to Week 4 was 0.33 points between the NT 201 and Botox groups. In the primary efficacy analysis, change from Baseline in TWSTRS-Severity score at Week 4, NT 201 was non-inferior to Botox, because the upper 95% confidence bound of the mean difference between the treatment effects was lower than the predefined difference Δ=1.3 points (Figure 5).

For the confirmatory statistical analysis of the primary efficacy variable, no missing data were imputed. However, the efficacy findings from the two specific strategies of adjustments for missing values (as stated earlier) did not differ from the findings obtained from the TPP sample.

Figure 5: Non-Inferiority (LSMean NT 201 - LSMean BTXCo) Based on the Final ANCOVA Model for Change from Baseline in the TWSTRS-Severity Score (TPP)



Source: Study report

TWSTRS - Pain Subscore

The two-sample Wilcoxon-test finds no significant differences between NT 201 and BTXCo in the change from baseline in the TWSTRS - Pain subscore observed at week 4 visit (p=0.4076).

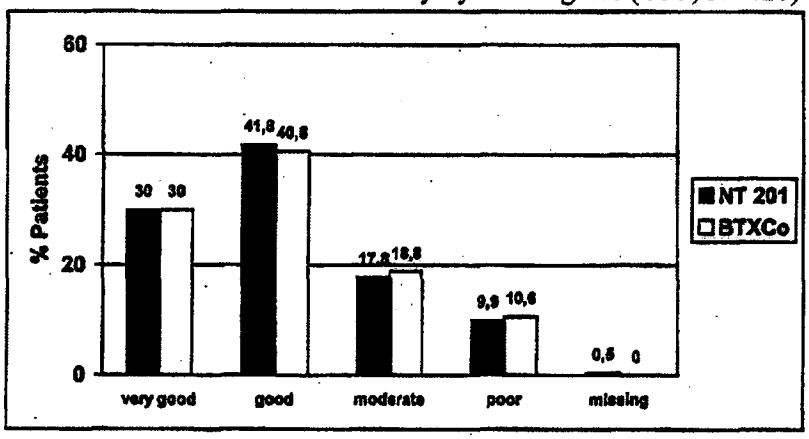
VAS for Pain

The two-sample Wilcoxon-test also finds no significant differences between the NT 201 and BTXCo treatment groups in the change from baseline in the VAS pain score observed at week 4 visit (p=0.2892).

Investigator's Global Assessment of Efficacy

The two-sample Wilcoxon test reveals a p-value of 0.8000 for the difference between the treatment groups with respect to the investigator's global assessment of efficacy (Figure 6).

Figure 6: Global Assessment of Efficacy by Investigator (TPP, N=420)



Source: Study report

The positive results of the secondary efficacy endpoint analyses also supported the primary efficacy finding. NT 201 was similar to Botox in the reduction of TWSTRS-Severity Score, and in all secondary efficacy parameters.

Study 0003 (BEB)- Supportive study

Study 0003 was a Phase 3, double blind, randomized, non-inferiority, active-controlled, multicenter study. Adult subjects with bilateral BEB and a stable clinical response to the two most recent previous injections of Botox were included in the study. The study was conducted at 42 centers in Europe and Israel.

Treatments:

There were 300 subjects in the ITT Population (148 in the NT 201 group and 152 in the BTXCo group- registered in Europe as BOTOX) and 256 in the TPP Population. A total of 294 subjects completed the study. Subjects in the NT 201 group received a mean total dose (both eyes) of 41 U, and subjects in the Botox group received a mean total dose of 42 U.

Primary Efficacy Measure

The primary efficacy variable was the change in mean JRS sumscore from Baseline to Week 3. The primary efficacy variable was analyzed in the treated per-protocol (TPP) population. Analysis of covariance (ANCOVA) was used to test the efficacy of the primary variable with the change in the JRS sum score as the dependent variable and with at least the JRS sum score at baseline and treatment group as independent variables. Other covariates included in this model were total dose, sex, age, number of injection sessions since diagnosis of blepharospasm (grouped as 0-2 sessions, 3-5 sessions and >5 sessions), pooled country, and the treatment*pooled country interaction. The final model used for statistical inference included all variables having an influence on the primary efficacy variable of $p<0.2$. Backward selection was used for model building purposes and both the full and final models were incorporated in this study.

The mean difference of the changes of the JRS sum score from baseline was calculated from the final ANCOVA model as difference of the least square means of the change (NT 201) minus change (BTXCo). The change was defined as the value at three weeks minus the value at baseline. NT 201 was considered clinically non-inferior if the upper 95% confidence bound (UCB) was less than Δ. Δ was defined as 0.8 points of the JRS. If the UCB was <0, NT 201 was considered superior to BTXCo if also confirmed in the ITT population. The one-sided significance level was set to $\alpha=0.025$.

The influences of missing data values on this primary efficacy endpoint were explored using two strategies (i) all missing values at the control and final visits in both treatment groups were set to baseline values, (ii) missing values were replaced by the mean value of the corresponding treatment group at the corresponding visit.

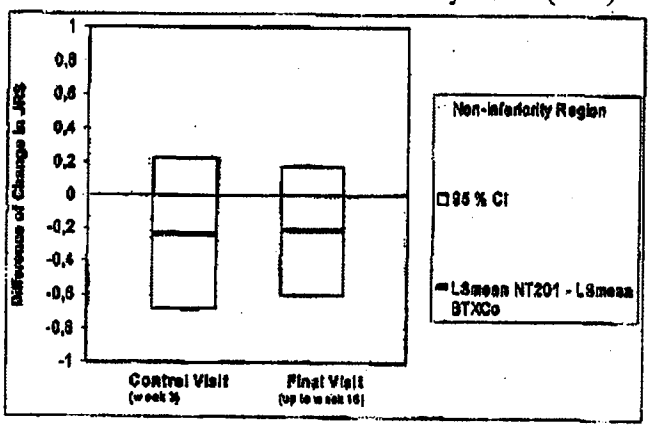
Secondary Efficacy Variables:

The secondary efficacy measures were (i) Change from baseline in the JRS sum score at the final visit, (ii) Change from baseline in the mean total score for the Function Scale for Patients with Blepharospasm (BSDI) at the control and final visits, (iii) Assessment of efficacy by the investigator, and (iv) Duration of treatment effect. ANCOVA models and two sample Wilcoxon Rank-Sum test were used to analyze the secondary measures.

Efficacy Findings

The Lsmean difference in the JRS sum scores between treatments was -0.23 with the 95% confidence interval between -0.68 and 0.22. The upper confidence bound UCB of the 95% CI was less than 0.8 (figure 7). Therefore, NT 201 can be considered clinically non-inferior to BTXCo in the treatment of blepharospasm. This result was confirmed in the analysis of the ITT population.

Figure 7: Non-Inferiority (LSMean NT 201 - LSMean BTXCo) Based on the Final ANCOVA Model for Change from Baseline in the TWSTRS-Severity Score (TPP)



Source: Study report

Change from Baseline in JRS Sum Score at the Final Visit

Both treatments groups showed reductions (i.e., an improvement in blepharospasm symptoms) in the mean sum scores of the JRS at the control and final visits. This improvement was slightly better in patients in the NT 201 group compared to those in the BTXCo group.

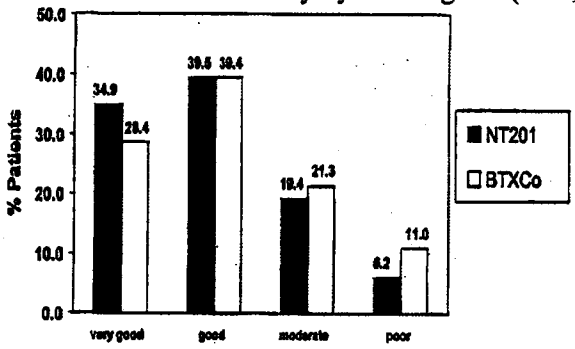
Change from Baseline in Mean Total Score of the BSDI at the Control and Final Visits

The change from baseline (control visit: -0.83 NT 201 and -0.82 BTXCo , final visit: -0.36 NT 201 and -0.22 BTXCo) indicates that patients in the NT 201 group showed a trend towards a greater improvement in blepharospasm at both the control and final visits than those in the BTXCo group.

Investigator's Global Assessment of Efficacy

The two-sample Wilcoxon test reveals a p-value of 0.14 for the difference between the treatment groups with respect to the investigator's global assessment of efficacy (Figure 8).

Figure 8: Global Assessment of Efficacy by Investigator (TPP, N=420)



Source: Study report

The positive results of the secondary efficacy endpoint analyses also supported the primary efficacy finding. NT 201 was similar to Botox in reduction of the JRS sum score from baseline, and in all secondary efficacy parameters.

3.1. FDA Reviewer's Data Analyses and Comment

This reviewer re-analyzed the efficacy data of the pivotal and supportive studies according to the protocol specified statistical analysis plans and found that the statistical findings are consistent with the sponsor's reported efficacy findings. In each study, a few patients dropped out from the study before the protocol defined primary study endpoints, and hence the missing data had no impact on the efficacy conclusions of the studies.

4. Subgroup Analysis

Subgroup analyses were performed for the primary efficacy measure at week 6 by age (<60, ≥60), and gender for the two pivotal studies. Both studies were conducted in North America. For the primary efficacy variables of the two studies, there were no differences in efficacy of NT 201 by age and gender subgroups at week 6 (Tables 7 & 8).

Table 7: Subgroup Analyses of the Primary Efficacy Variable (Mean Change (SD) in TWSTRS-Total Score, Baseline to Week 4) in Study 0408/1 (CD) Main Period (points; Missing Values Replaced by Patient's Baseline Value; ITT Population)

	240 U NT 201 (N=81)		120 U NT 201 (N=78)		Placebo (N=74)	
	Mean Change (SD)		Mean Change (SD)		Mean Change (SD)	
Female	(n=54)	-12.61 (12.77)	(n=51)	-10.13 (10.11)	(n=49)	-2.92 (8.34)
Male	(n=27)	-7.46 (8.46)	(n=27)	-9.40 (10.97)	(n=25)	-0.81 (4.50)
Age (years)						
≤60	(n=55)	-10.64 (12.09)	(n=57)	-10.11 (10.77)	(n=58)	-2.32 (7.28)
>60	(n=26)	-12.32 (11.07)	(n=21)	-9.72 (9.40)	(n=16)	-2.06 (8.01)

Source: study report

Table 8. Subgroup Analysis: Mean (±SD) Change from Baseline to Week 6 in JRS Severity Subscore in Study 0433/1 (BEB) Main Period (Missing Values Replaced by Last Observation Carried Forward; ITT Population)

	NT 201(N=75)		Placebo (N=34)	
	Mean Change (SD)		Mean Change (SD)	
Female	(n=49)	- 1.0 (1.23)	(n=22)	0.2 (0.80) 0.3
Male	(n=26)	- 0.5 (0.99)	(n=12)	(1.14)
Age (years)				
≤60 years	(n=35)	- 0.8 (1.12)	(n=14)	0.4 (0.65)
>60 years	(n=40)	- 0.8 (1.24)	(n=20)	0.1 (1.05)

Source: study report

5. Summary and Conclusions

Xeomin has demonstrated its efficacy in treating patients with cervical dystonia. In the study 408/1, the primary efficacy endpoint was the change in the TWSTRS total score from baseline to Week 4 post injection. In the ITT population, the difference between the Xeomin 240 U group and the placebo group in the change of the TWSTRS total score from baseline to Week 4 was -9.0 points; the difference between the Xeomin 120 U group and the placebo group in the change of the TWSTRS total score from baseline to Week 4 was -7.5 points. Comparison of each Xeomin group to the placebo group was statistically significant at $p<0.001$. Subgroup analyses also revealed that the efficacy of the two Xeomin doses was similar in pre-treated subjects (i.e., subjects who had received a Botulinum toxin prior to this study) and in naïve subjects (i.e., subjects who had not received a Botulinum toxin prior to this study). Examination of age and gender subgroups did not identify differences in response to Xeomin among these subgroups. There were a few African-American subjects to adequately assess efficacy in that population.

Xeomin also has demonstrated its efficacy in treating patients with benign essential blepharospasm. In the study 433/1, the highest dose permitted was 50 U per eye; the mean Xeomin dose was approximately 33 U per eye. The primary efficacy endpoint was the change in the JRS Severity subscore from baseline to Week 6 post injection. In the ITT population, the difference between the Xeomin group and the placebo group in the change of the JRS Severity subscore from baseline to Week 6 was -1.0 points. Comparison of the Xeomin group to the placebo group was statistically significant at $p<0.001$. Examination of age and gender subgroups did not identify substantial differences in response to Xeomin among these subgroups.