



Australian Government
Department of Health
Therapeutic Goods Administration

AusPAR Attachment 2

Extract from the Clinical Evaluation Report for Botulinum toxin type A

Proprietary Product Name: Xeomin

Sponsor: Merz Australia Pty Ltd

Date of CER: 20 May 2013

About the Therapeutic Goods Administration (TGA)

- The Therapeutic Goods Administration (TGA) is part of the Australian Government Department of Health, and is responsible for regulating medicines and medical devices.
- The TGA administers the *Therapeutic Goods Act 1989* (the Act), applying a risk management approach designed to ensure therapeutic goods supplied in Australia meet acceptable standards of quality, safety and efficacy (performance), when necessary.
- The work of the TGA is based on applying scientific and clinical expertise to decision-making, to ensure that the benefits to consumers outweigh any risks associated with the use of medicines and medical devices.
- The TGA relies on the public, healthcare professionals and industry to report problems with medicines or medical devices. TGA investigates reports received by it to determine any necessary regulatory action.
- To report a problem with a medicine or medical device, please see the information on the TGA website < <http://www.tga.gov.au>>.

About the Extract from the Clinical Evaluation Report

- This document provides a more detailed evaluation of the clinical findings, extracted from the Clinical Evaluation Report (CER) prepared by the TGA. This extract does not include sections from the CER regarding product documentation or post market activities.
- The words (Information redacted), where they appear in this document, indicate that confidential information has been deleted.
- For the most recent Product Information (PI), please refer to the TGA website < <http://www.tga.gov.au/hp/information-medicines-pi.htm>>.

Copyright

© Commonwealth of Australia 2014. This work is copyright. You may reproduce the whole or part of this work in unaltered form for your own personal use or, if you are part of an organisation, for internal use within your organisation, but only if you or your organisation do not use the reproduction for any commercial purpose and retain this copyright notice and all disclaimer notices as part of that reproduction. Apart from rights to use as permitted by the *Copyright Act 1968* or allowed by this copyright notice, all other rights are reserved and you are not allowed to reproduce the whole or any part of this work in any way (electronic or otherwise) without first being given specific written permission from the Commonwealth to do so. Requests and inquiries concerning reproduction and rights are to be sent to the TGA Copyright Officer, Therapeutic Goods Administration, PO Box 100, Woden ACT 2606 or emailed to < tga.copyright@tga.gov.au>.

Contents

List of abbreviations	5
1. Introduction	8
2. Clinical rationale	8
3. Contents of the clinical dossier	9
3.1. Scope of the clinical dossier	9
3.2. Paediatric data	9
3.3. Good clinical practice	9
4. Pharmacokinetics	10
5. Pharmacodynamics	10
5.1. Studies providing pharmacodynamic data	10
5.2. Summary of pharmacodynamics	11
5.3. Evaluator's overall conclusions on pharmacodynamics	17
6. Dosage selection for the pivotal studies	17
7. Clinical efficacy	17
7.1. Pivotal efficacy studies	17
7.2. Glabellar Frown Lines (GFL)	17
7.3. Cervical Dystonia	37
7.4. Spasticity	54
7.5. Supportive efficacy studies	62
7.6. Analyses performed across trials	73
7.7. Dosing considerations	74
7.8. Blepharospasm	75
7.9. Evaluator's conclusions on clinical efficacy	83
8. Clinical safety	84
8.1. Studies providing evaluable safety data	84
8.2. Patient exposure	85
8.3. Adverse events	85
8.4. Treatment-related adverse events (adverse drug reactions)	87
8.5. Serious adverse events	87
8.6. Deaths	88
8.7. Discontinuation due to adverse events	88
8.8. Adverse events of special interest	89
8.9. Immunogenicity	89
8.10. Laboratory tests	90

8.11.	Liver function _____	91
8.12.	Kidney function _____	91
8.13.	Other clinical chemistry _____	91
8.14.	Haematology _____	91
8.15.	Electrocardiograph _____	92
8.16.	Vital signs _____	92
8.17.	Post-marketing experience _____	93
8.18.	Safety issues with the potential for major regulatory impact _____	93
8.19.	Other safety issues _____	93
8.20.	Evaluator’s overall conclusions on clinical safety _____	94
9.	First round benefit-risk assessment _____	94
9.1.	First round assessment of benefits _____	94
9.2.	First round assessment of risks _____	95
9.3.	First round assessment of benefit-risk balance _____	95
10.	First round recommendation regarding authorisation _____	95
11.	Clinical questions _____	95
11.1.	Pharmacokinetics _____	95
11.2.	Pharmacodynamics _____	95
11.3.	Efficacy _____	95
11.4.	Safety _____	95
12.	References _____	95

List of abbreviations

Abbreviation	Meaning
AB	Antibody
ADL	Activity of daily living
AE	Adverse event
AESI	Adverse event of special interest
ANCOVA	Analysis of covariance
BLEPH	Blepharospasm
BfArM	Bundesinstitut für Arzneimittel und Medizinprodukte (German Federal Institute for Drugs and Medical Devices)
BoNT/A	Botulinum neurotoxin type A
BTX-A	Botulinum toxin type A
BMI	Body mass index
BSDI	Blepharospasm disability index
CAmR	CMAP M-wave amplitude reduction
CETS	Composite endpoint treatment success
CD	Cervical dystonia
CI	Confidence interval
CMAP	Compound muscle action potential
CSR	Clinical study report
DAS	Disability assessment scale
ECG	Electrocardiogram
EDB	Extensor digitorum brevis
EEA	European Economic Area
EFS	Evaluable for safety
E _{max}	Maximum effect
EP	Extension period

Abbreviation	Meaning
EU	European Union
FAS	Full analysis set
FAT	Frenchay arm test
FDA	US Food and Drug Administration
FIA-AB	Fluorescence immunoassay for antibodies
FWS	Facial wrinkle scale
GATR	Global assessment of treatment response
GFL	Glabellar frown lines
HDA	Mouse ex vivo hemidiaphragm assay
HV	Healthy volunteer
IM	Intramuscular
ITT	Intention-to-treat
IVRS	Interactive voice response system
JRS	Jankovic Rating Scale
kD	kiloDalton
LOCF	Last observation carried forward
LS mean	Least square mean
MRZ	Merz
MP	Main Period
n	number of observed cases
N	Number in group
NA	Not applicable
OR	Odds ratio
OLEX	Open-label extension
PEGR	Patient evaluation of global response

Abbreviation	Meaning
PROM	Passive range of motion
SCM	Sternocleidomastoid muscle
SD	Standard deviation
SP	Spasticity
SPL	Splenius capitis muscle
T _{max}	Time to maximum effect
TPP	Treated-per-protocol
TTV	Trial termination visit
TWSTRS	Toronto Western Spasmodic Torticollis Rating Scale
U	Units
USA	United States of America
VAS	Visual analogue scale
WE MOVE	Worldwide Education and Awareness of Movement Disorders

1. Introduction

NT 201 (Xeomin) is a member of the class of drugs referred to as botulinum toxins, of which the most well-known is Botox. These toxins are produced in nature by the bacterium, *Clostridium botulinum*, and contribute to the clinical paralytic syndrome of botulism that sometimes results from clostridial infection. The agents have been adopted for topical therapeutic use in situations where local muscle weakness is desirable, such as conditions characterised by spasticity, dystonia or other muscle hyperactivity.

NT 201 differs from Botox and other members of this class in that it lacks the complexing proteins that normally accompany the toxin.

2. Clinical rationale

For a range of neuromuscular conditions, overactivity of muscles is a major contributor to disability, and botulinum toxin preparations have been useful in ameliorating symptoms. Such conditions include spasticity and dystonia, for which there are only limited therapeutic options.

Anti-spasticity drugs include baclofen and some anticonvulsants such as valproate or benzodiazepines, but these agents are usually administered systemically and often produce unacceptable drowsiness. Some dystonias respond to levodopa or dopamine agonists, but the response is often unsatisfactory.

Cervical dystonia in which patients experience unwanted, involuntary head turning is an uncomfortable condition that may impair motor function and produce social stigma and isolation.

Blepharospasm, which is usually considered a dystonia of the muscles closing the eye, is also disfiguring and socially isolating, but can in addition produce functional blindness if patients are unable to open their eyes.

Botulinum toxin treatment provides a useful alternative to systemic treatment for these conditions, because it is injected into the target muscles and has few or no systemic side effects. It has therefore become a widely accepted treatment, usually administered in specialised clinics by neurologists or rehabilitation physicians.

Botulinum toxin has also been used to treat facial wrinkles. There is no strong clinical rationale for such treatment, but there has been extensive experience with this approach and the overall safety profile of existing botulinum toxins when used for cosmetic purposes is acceptable.

The new agent NT 201 does not offer any major changes in efficacy or safety over existing botulinum agents, but it is a simpler preparation than Botox because it lacks the complexing proteins that normally accompany the bacterial toxin. These proteins do not contribute to the functional effects of botulinum toxin, and dissociate from the toxin soon after injection, but the complexing proteins are likely to contribute to the overall antigenic stimulus associated with the injection of foreign protein. NT 201 would therefore be expected to be less immunogenic than complexed forms of the toxin.

The sponsor has submitted some evidence, based on animal studies, supporting this idea; these studies show that neutralising antibodies are produced less commonly after NT 201 administration than after other botulinum toxin preparations.

Although NT 201 is proposed as a new chemical entity, it is equivalent to the active agent of existing products, the clinical rationale for which is already well-established.

3. Contents of the clinical dossier

3.1. Scope of the clinical dossier

The application to register NT 201 is primarily based on eight pivotal Phase III studies, five of which were double-blind and placebo-controlled (Study 0724 (GFL), Study 0741 (GFL), Study 0408/1 (CD), Study 0433/1 (BLEPH), and Study 0410/i (SP)).

In addition, the sponsor submitted two non-inferiority double-blind, active-controlled studies in comparison with Botox (Study 0013/1 (CD), Study 0003/1 (BLEPH)) and one controlled observer-blind, parallel-group Phase III study with two different dilutions of NT 201 (Study 0607/1 (SP)).

The contents of the submission are summarised below:

- Three clinical pharmacology studies, all of which provided pharmacodynamic (PD) data.
- No pharmacokinetic (PK) studies, which is appropriate given the toxic nature of NT 201 and its topical application.
- Eight pivotal Phase III trials (0724, and 0741 (GFL), 0408 and 0013 (CD), 0433 and 0003 (BLEPH), 0410 and 0607 (SP)).¹
- Several supportive efficacy studies, including Study 0527 (GFL), a randomised, double-blind, placebo-controlled, Phase II dose finding study, Study 520 (GFL), Study GL 3001 (GFL), which was uncontrolled, Study GL 3002 (GFL) in comparison with Vistabel, Study 9801 (CD) in comparison with Botox, Study 0013/1 (CD) in comparison with Botox, Study 0003/1 (BLEPH), in comparison with Botox and Study 0607/1 (SP), which assessed different dilutions of NT 201.
- Repeated-dose studies, including pooled long-term, repeated-dose safety data from studies with multiple administrations of NT 201: Study 0609 and its lead-in Studies 0520, 0527, 0724, and 0741 for the GFL indication; Study 0408/1 (Main Period) and 0408/2 (Extension Period), and Study 0605/1 (Main and OLEX Period) for the CD indication; Study 0433/1 (Main Period) and 0433/2 (OLEX Period) for the BLEPH indication; and Study 0410/1 (Main Period) and 0410/2 (OLEX Period) for the SP indication.
- An Integrated Summary of Efficacy and an Integrated Summary of Safety.
- The sponsor's clinical overview, summary of clinical efficacy, summary of clinical safety and literature references.

3.2. Paediatric data

The submission did not include paediatric data.

3.3. Good clinical practice

For each submitted study, a declaration was provided indicating that the study had conformed to Good Clinical Practice guidelines.

¹ In the study summaries, the sponsor abbreviated the study names to improve readability, and this evaluation report follows the same conventions. The prefixes 'MRZ 60201-' and 'BTC-60201' have been removed from the titles of individual studies, and the indication being studied has been added in parentheses (that is, glabellar frown lines (GFL), cervical dystonia (CD), blepharospasm (BLEPH), spasticity (SP), or healthy volunteer (HV)). The suffix "/1" indicates the main period of a study, and "/2" indicates the extension period.

4. Pharmacokinetics

Conventional PK studies cannot be performed with botulinum toxin in any of its forms, because it is extremely toxic when administered systemically. It is always administered locally, in very small quantities (picograms of toxin per injection), directly at the intended site of action or within a short distance of the intended target; it then spreads over a limited distance by local diffusion. The NT 201 preparation is no different to existing botulinum toxin preparations in this regard.

Even though NT 201 represents a new preparation of botulinum toxin, it is expected to have similar diffusion characteristics to standard botulinum toxin preparations. Botox, for instance, consists of the same neurotoxin as NT 201, with an identical amino-acid sequence, but is associated with complexing proteins. These proteins dissociate from Botox and other marketed forms of the neurotoxin almost immediately after intramuscular injection; from that step onwards, the diffusing properties of the different preparations would be expected to be similar.

Some measure of the diffusing kinetics of NT 201 was provided by the PD study, Study 0709, in which NT 201 was injected into the forehead and the subsequent area of anhidrosis assessed. In this study, the diffusion of NT 201 appeared similar to the competing products Vistabel and Dysport.

Overall, no new PK issues are raised by the proposed introduction of NT 201.

5. Pharmacodynamics

5.1. Studies providing pharmacodynamic data

The sponsor submitted three PD studies. The main conclusions of the individual PD studies are discussed below.

Two of the Studies (9901 and 0113) investigated the effect of NT 201 on muscle activity in healthy volunteers, as measured by the compound motor action potential (CMAP) in comparison to Botox.

The third Study (0709) assessed the effect of NT 201 on sweating, by measuring the area of anhidrosis following an injection, in comparison to the commercial products Vistabel and Dysport; this is primarily an assessment of the different drugs' diffusion characteristics.

Table 1: Summary of completed Phase I studies for NT 201 in neurology

Trial Countries (Study report location)	Phase, Population and Design	Study Treatment	Primary Endpoint
9901 (HV) Germany (Section 5.3.4.1)	Phase 1 14 healthy male volunteers Open-label, within-treatment comparison	4 U NT 201 or Botox intramuscular	Change in maximal amplitude of CMAP
0113 (HV) Germany (Section 5.3.4.1)	Phase 1b 32 healthy male volunteers Double-blind, within-treatment comparison examining dose response profile, systemic diffusion and duration of effect	2, 4, 16, or 32 U of NT 201 or Botox intramuscular	Reduction in CMAP M-wave at Week 4. Follow up for 52 weeks
0709 (HV) Germany (Section 5.3.4.1)	Phase 1 29 healthy female volunteers Double-blind, randomized, single-dose, single-center study	5 U NT 201 or 5 U Vistabel or 12.5 U Dysport intramuscular	Maximal area of anhidrosis

CMAP: Compound muscle action potential; HV: healthy volunteer (subject); U: Units

None of the PD studies had deficiencies that excluded their results from consideration.

5.2. Summary of pharmacodynamics

5.2.1. Mechanism of action

The mechanism of action of botulinum toxin A has been studied for decades, principally using existing commercial botulinum toxin preparations, such as Botox, that include complexing proteins. The mechanism of action of NT 201 is expected to be identical to that of Botox, particularly because Botox has been shown to dissociate from its complexing proteins soon after injection, releasing the toxin.²

The sponsor summarises this as follows: *'The Botulinum toxin complex (900 kDa) dissociates largely upon reconstitution and dilution in the vial and completely upon IM injection into the neutral pH tissue environment.² Accordingly, there is no difference between drug products with and those without complexing proteins after IM injection and there is no difference in diffusion behaviour.'*

That is, Botox can be considered a pro-drug of NT 201.

The sponsor notes that, when botulinum toxin preparations are administered orally to mice, there do seem to be important differences between preparations with and without the complexing proteins. In particular, the LD50 was very high for oral NT 201 (10,000 fold higher than via the intravenous route, and higher than for other botulinum toxins), suggesting that the complexing proteins normally act to preserve the oral potency of clostridial toxin, but these complexing proteins do not exert any significant PD effect for intramuscular toxin.

From studies in Botox and similar preparations, it is known that clostridium botulinum neurotoxin type A blocks the presynaptic release of acetylcholine from the neuromuscular endplates of peripheral nerves, which prevents the transmission of nerve impulses to the muscle in the form of acetylcholine.

A four step process has been identified.

- The first step involves high-affinity binding of the neurotoxin to cell surface receptors on the presynaptic membrane. The heavy chain of the neurotoxin binds to gangliosides (GT1b) and to the synaptic vesicle protein SV2 (isoform C).³
- After binding, the neurotoxin is transferred via endocytosis into an endocytotic vesicle, or endosome.
- The light chain of the neurotoxin then enters the cytosol, with the help of the translocation domain of the heavy chain.⁴
- The light chain acts as a zinc-dependent protease, cleaving synaptosomal-associated protein (SNAP)-25, a protein involved in the targeting and fusion of synaptic vesicles with the presynaptic membrane, important preliminary steps prior to transmitter release. By cleaving SNAP-25, the toxin interferes with the release of acetylcholine from nerve terminals, equivalent to chemical denervation of the muscle.⁵

Secondary effects, resulting from the denervation, include degeneration of the neuromuscular junctions, cessation of muscle activity and reversal of muscle hypertrophy.

² Eisele, 2011.

³ Dong M, Yeh F, Tepp WH, et al. 2006. SV2 is the protein receptor for botulinum neurotoxin A. *Science* 312(5773):592-596; Mahrhold S, Rummel A, Bigalke H, et al. 2006. The synaptic vesicle protein 2C mediates the uptake of botulinum neurotoxin A into phrenic nerves. *FEBS Lett*;580(8):2011-2014.

⁴ Brunger AT, Breidenbach MA, Jin R, et al. 2007. Botulinum neurotoxin heavy chain belt as an intramolecular chaperone for the light chain. *PLoS Pathog* 3(9):1191-1194.

⁵ Schiavo G, Montecucco C. 1997. The Structure and Mode of Action of Botulinum and Tetanus Toxins. In: *The Clostridia: Molecular Biology and Pathogenesis*. (18):295-322. Academic Press Limited.

Dolly JO. 1997. Therapeutics and research exploitation of botulinum neurotoxins. *Eur J Neurol* 4(Suppl 2):S5-S10.

The effects of botulinum toxin are largely reversible, though the recovery process often takes months and may continue even longer.⁶ During the recovery phase, growth factors are released, nerve terminals and motor endplates are re-established, and axons sprout collaterals.⁷ Recovery of impulse transmission begins after approximately eight to 24 weeks, and typically takes approximately 12 weeks. Eventually, the collaterals regress,^{8,9} and normalisation of morphology and function occurs.

There is also evidence that botulinum toxin can reduce muscle spindle afferent discharge by inhibiting the gamma motor neuron cholinergic junction on the intrafusal fiber.¹⁰ Botulinum neurotoxin also blocks local acetylcholine release in the autonomic nervous system, including eccrine (sweat) glands, which means it can be used to treat hyperhidrosis. Salivary glands are also susceptible to denervation, as evidenced by side effects such as a dry mouth when patients received botulinum toxin for treatment of cervical dystonia.⁹

Because NT 201 contains the same active toxin as Botox, its mechanism of action is thought to be identical, primarily mediated by cleavage of the SNAP-25 protein, and subsequent inhibition of acetylcholine release. No clinical studies were submitted clarified this mechanism of action but, given the close similarities between the agents, this is acceptable. The sponsor provided evidence, discussed below, that NT 201 has very similar potency to Botox.

5.2.2. Pharmacodynamic effects

5.2.2.1. Primary pharmacodynamic effects

Two studies were submitted that involved a comparison of the primary pharmacodynamics of NT 201 and Botox in healthy volunteers. The paralysing effect on muscle was quantified with the compound motor action potential (CMAP), recorded with standard neurophysiological equipment. Within the statistical power limitations of the studies, the two different botulinum toxin preparations showed a similar effect on muscle in both studies, broadly indicating dose equivalence.

In Study 9901 (n = 14), male volunteers aged 20 to 45 years received a single intramuscular (IM) injection of 4 U of NT 201 or 4 U of Botox into the extensor digitorum brevis (EDB) muscle of the right and left foot. Subjects were randomised to one of two groups - seven subjects receiving NT 201 on the left, seven subjects on the right, with Botox administered to the contralateral foot - so subjects served as their own controls.

Sample size estimation indicated a minimum of 12 EDB muscles for each treatment, based on data from a pilot trial with 10 healthy volunteers. Given an alpha level of 0.05, a one-sided border of irrelevant inferiority of NT 201 versus Botox of 15%, the study was intended to have a power of 76% to show equivalence in duration of effect.

The magnitude of the muscle paralysis was similar for each drug, and exhibited a similar time course over the 90 days following injection. NT 201 was associated with slightly greater potency overall, but this was not statistically significant and at times the effect was marginally more potent with Botox, (Most subjects showed a difference in the two feet, with stronger CMAPS on the right. This effect was greater than the overall difference between the two drugs, making firm

⁶ Herrera AA. 1997. Remodeling and Elimination of Neuromuscular Synapses. The Synapse in Development, Health, and Disease. *Advances in Organ Biology* (2):29-53. Greenwich CT, JAI Press Inc. Festoff BW, Hantai D, Citron B.

⁷ De Paiva A, Meunier FA, Molgo J, et al. 1999. Functional repair of motor endplates after botulinum neurotoxin type A poisoning: Biphasic switch of synaptic activity between nerve sprouts and their parent terminals. *Proc Natl Acad Sci U S A* 96(6):3200-3205.

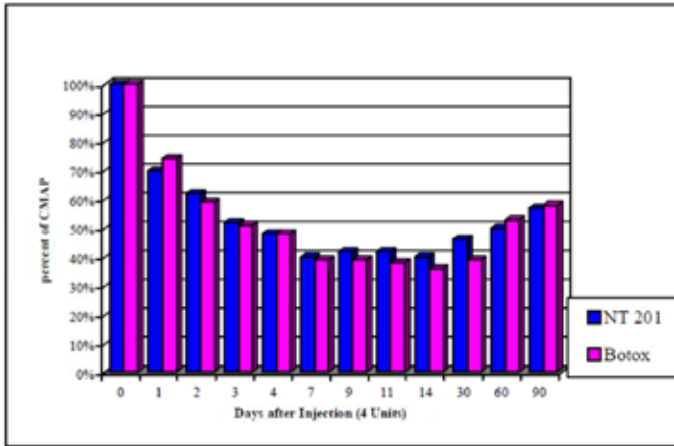
⁸ Borodic GE, Joseph M, Fay L, et al. 1990. Botulinum A toxin for the treatment of spasmodic torticollis: dysphagia and regional toxin spread. *Head Neck* 12(5):392-399.

⁹ Kessler KR, Benecke R. 1997. Botulinum toxin: From poison to remedy. *Neurotoxicology* 18(3):761-770.

¹⁰ Filippi GM, Errico P, Santarelli R, et al. 1993. Botulinum A toxin effects on rat jaw muscle spindles. *Acta Otolaryngol* 113(3):400-404.

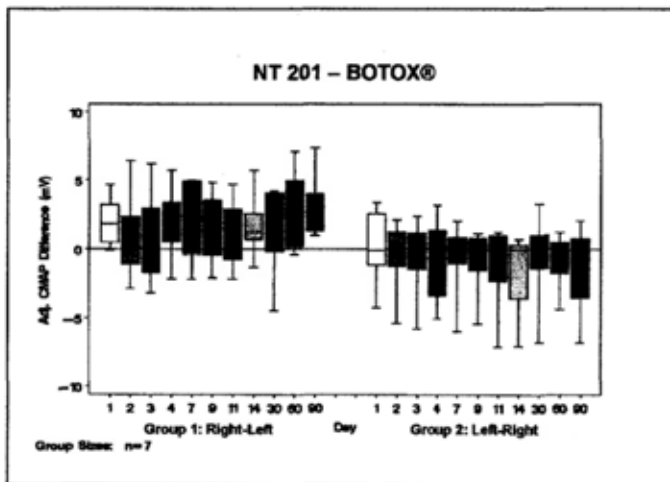
conclusions difficult; the second figure below shows the between-treatment difference adjusted for the right-left difference). Onset of effect was defined as a reduction of CMAP to less than or equal to 70% of individual baseline, and most subjects showed an onset of effect with both treatments one day after injection.

Figure 1: Effect of NT 201 and Botox on CMAP as measured by electromyography after supramaximal electrical stimulation (N = 14)

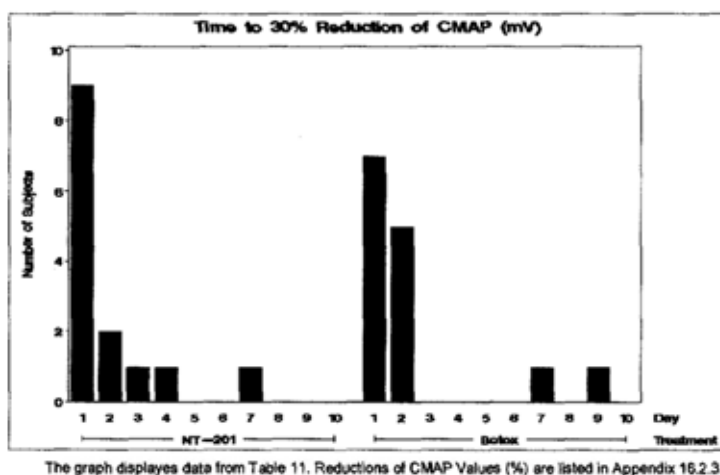


CMAP: Compound muscle action potential

Figure 2: Adjusted CMAP difference (mV) between NT 201 and Botox



The box plots show the median, quartiles, min. and max. CMAP values.
 (The first-named side is the side that received NT-201)†

Figure 3: Time to 30% reduction (days) of CMAP values

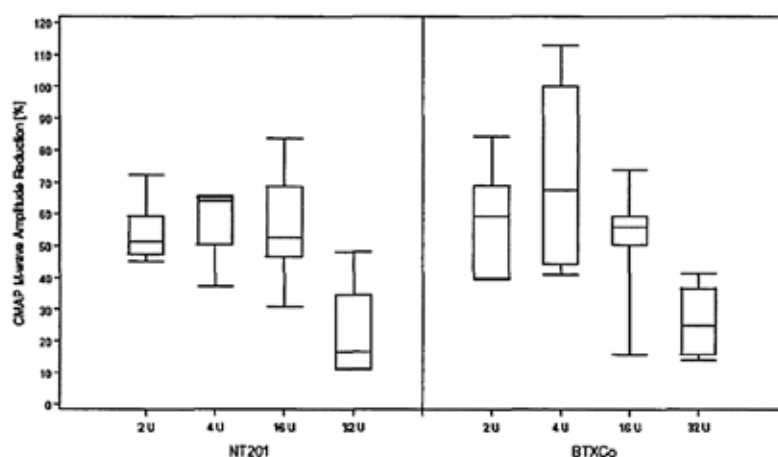
In Study 0113, volunteers (evaluable $n = 20$) received IM injections of two, four, 16 or 32 U of NT 201 in the EDB muscle of one foot, and the same dose of Botox in the other foot. Dose and foot allocation were randomised. The magnitude of the paralysis was comparable for NT 201 and Botox when matched for dose, but there were several flaws in the randomisation process that render the results somewhat unreliable. The study was also underpowered. A sensitivity analysis showed that seven volunteers per dose group were needed. Anticipating one to two dropouts in at least one dose group, a sample size of eight volunteers per dose group was chosen, giving a total sample size of $N = 32$. Of 32 randomised subjects, however, only 26 had efficacy data and, of these, six had to be rejected because of randomisation errors, so only 20 were considered evaluable.

Table 2: CMAP M-wave amplitude reduction (%) at Week 4 in accurately randomised volunteers (ITT ARV)

	Dose Group				Total N=20
	2 U N=6	4 U N=4	16 U N=6	32 U N=4	
	Mean \pm SD				
	Geometric Mean, Coefficient of Variation				
	Median (Range)				
	p-Value (95%-Confidence Interval)				
NT 201	54.5 \pm 10.30 53.8, 0.18 51.5 (45.1, 72.5) <0.001 (44.6, 64.9)	57.9 \pm 13.65 56.4, 0.28 64.2 (37.5, 65.7) 0.025 (36.5, 87.2)	55.9 \pm 18.30 53.3, 0.35 52.5 (31.0, 83.8) 0.006 (37.3, 76.3)	23.2 \pm 17.31 19.2, 0.77 16.6 (11.2, 48.1) 0.017 (6.5, 56.7)	49.3 \pm 19.36 44.1, 0.60 50.0 (11.2, 83.8) <0.001 (34.0, 57.1)
BTXCo	58.6 \pm 17.54 56.4, 0.31 59.2 (39.7, 84.4) 0.006 (41.0, 77.7)	72.3 \pm 33.92 66.4, 0.51 67.6 (41.1, 113.0) 0.187 (30.9, 142.8)	52.1 \pm 19.32 47.3, 0.59 56.2 (16.1, 73.9) 0.020 (26.7, 83.8)	26.3 \pm 12.59 24.1, 0.53 25.1 (14.1, 41.2) 0.011 (10.9, 53.2)	52.9 \pm 24.99 46.6, 0.60 52.3 (14.1, 113.0) <0.001 (36.0, 60.5)

Data source: Table 14.5.2.1.1.1

Figure 4: Boxplot of the CMAP M-wave amplitude reduction (%) at Week 4 in comparison to baseline (100%) in accurately randomised volunteers (ITT ARV)



In the third PD study, Study 0709, the effects of three treatments on forehead sweating were compared in healthy female volunteers: NT 201, 5 U (incobotulinumtoxin A without complexing proteins); Vistabel, 5U (onabotulinumtoxin A with complexing proteins); and Dysport 12.5 U (abobotulinumtoxin A). Onabotulinumtoxin and abobotulinumtoxin are chemically distinct forms of the toxin, with slightly different molecular weights and different potencies. The higher dose of Dysport reflects general dosing recommendations, which suggest that 2.5 to 3.0 U of Dysport are equivalent to 1 unit of Botox (or Vistabel, which is essentially the same product as Botox but marketed for cosmetic usage).

Each subject received two of the three treatments, one on the right side of the forehead, and one on the left, as a single IM injection 3 cm above the orbital rim, directly above the pupil. Treatment was randomised and double-blind. A standardised sweating test (Minor, 192811) was performed, and documented with a standardised photo, then assessed by a blinded dermatologist. The maximal area of anhidrosis after six weeks was compared between the three treatments, using a linear model (with subject, treatment and side as initial factors) to produce point estimates and two-sided 90% confidence intervals. The analysis was considered exploratory, and there was no attempt to define formal equivalence criteria.

The basic results are shown below, with descriptive statistics, for the Per Protocol Set. Results in the Full Analysis Set were very similar (not shown). The subsequent table shows the 90%CI for the difference between each comparator and NT 201.

Table 3: Primary endpoint, maximal area of anhidrosis (mm²) within six weeks. Analysis population: Per-protocol-Set

Treatment	N	Mean	SD	CV	Minimum	Q1	Median	Q3	Maximum
NT201	13	364.3	138.13	37.9	220	270	298	430	630
Vistabel	16	343.1	110.72	32.3	199	228	345	426	540
Dysport	17	459.1	151.81	33.1	260	330	436	534	731

¹¹ Minor V. 1928. Ein neues Verfahren zu der klinischen Untersuchung der Schweißabsonderung. Deutsch Z Nervenhlk 101:302-306.

Table 4: Treatment contrasts for maximal area of anhidrosis (mm²) within 6 weeks. Lower and upper limits of 90% confidence intervals. Factors: subject, side and treatment. Analysis Population: Per-Protocol-Set

Treatment	LS Mean	SE	Change from NT201			
			Mean	SE	Lower	Upper
NT201	325.5	33.42				
Vistabel	373.7	31.52	48.2	30.97	-5.5	101.9
Dysport	459.2	31.06	133.7	30.23	81.3	186.1

The area of anhidrosis was similar after NT 201 and Vistabel, which were administered at the same dose (5 U). After Dysport, the area of anhidrosis was greater, indicating a greater area of diffusion in the forehead despite the fact that the Dysport dose (12.5 U) was intended to be comparable and the volume injected was identical. This could reflect the slightly lower molecular size of abobotulinumtoxin A compared to onabotulinumtoxin A.

The linear model broadly confirmed these findings. The confidence intervals included the possibility of no difference between NT 201 and Vistabel, but excluded zero for the comparison between NT 201 and Dysport, showing a significantly greater area of anhidrosis following Dysport.

Overall, this study suggests that NT 201 and Vistabel/Botox are similar in potency and diffusing characteristics, but the study was not powered to show formal equivalence. The deficiency is acceptable given that pivotal efficacy studies formally compared NT 201 and Botox, finding equivalence. The apparent difference between NT 201 and Dysport is of interest, but was not explored further in the study program. Differences between Dysport and NT 201 could pose potential problems for patients switching between therapies, particularly if clinicians forget to adjust the dose, but these problems already exist for patients switching between Botox and Dysport.

5.2.2.2. Secondary pharmacodynamic effects

Not applicable.

5.2.3. Time course of pharmacodynamic effects

The time course of paralysis following injection into the extensor digitorum brevis in Study 9901 is displayed above, in Section 4.2.2.1. The onset of an effect was observed in the CMAP within a day, and a substantial effect was still detectable after 90 days when the study concluded. The peak effect appeared to occur between seven and 14 days, but subjects were not assessed between Day 14 and Day 30, so this estimate is based on sparse sampling and may be unreliable.

The onset of detectable neurophysiological impairment does not necessarily correlate with the onset of a useful clinical effect (which might be dominated by residual, functioning nerve terminals rather than by the nerve terminals first showing a PD response). Also, for large muscles, some diffusion from the injection site to more distant nerve terminals is required before the onset of action so the time course is expected to change for different sites and indications. For glabellar frown lines, the toxin is injected under the skin and must diffuse into muscles before taking effect. Thus, in many subjects, the onset of action often takes longer than a day, despite the results of Study 9901. For a range of clinical conditions assessed in this submission, the subjective time to onset was noted by patients, and varied according to the condition. Time to offset also varied, but in most cases some benefit was still detectable after 12 weeks.

5.2.4. Relationship between drug concentration and pharmacodynamic effects

Given that NT 201 is applied locally and is not expected to produce detectable serum levels in the systemic circulation, knowledge about the relationship between concentration and effect is lacking.

5.2.5. Genetic-, gender- and age-related differences in pharmacodynamic response

No details were provided showing that individual responses to botulinum toxin vary according to genetic, gender or age-related factors. It is well known, from experience with other botulinum toxins, that individual patient responses are variable, but this issue was not explored in any PD studies. Much of this variability relates to the patient's pre-existing clinical condition, including the baseline muscle bulk and the degree of overactivity of the affected muscle. These factors would not be expected to pose new issues when using NT 201 compared to existing preparations.

5.2.6. Pharmacodynamic interactions

No new information was submitted in relation to PD interactions. NT 201 and other botulinum toxins should be used with caution in the setting of other muscle-weakening agents, including anti-spasmodic agents.

5.3. Evaluator's overall conclusions on pharmacodynamics

The submitted PD evidence was limited in scope, and one of the three submitted PD studies appeared unreliable because of a high number of protocol deviations.

On balance, the evidence suggests that NT 201 has a similar potency and similar time course of action compared to other botulinum toxin products, consistent with the notion that other agents including Botox dissociate to produce free toxin that is identical to NT 201.

6. Dosage selection for the pivotal studies

Given PD evidence suggesting dose equivalence between NT 201 and Botox, the doses used in pivotal studies were either based on treatment guidelines developed for Botox, in the case of treatment-naïve subjects, or consisted of the patients' previous Botox dose.

7. Clinical efficacy

7.1. Pivotal efficacy studies

The sponsor designated eight studies as pivotal – two studies for each indication. Many of the studies had an open-label extension, but these extensions are described separately. One of the studies designated as pivotal (0607) did not have a non-NT 201 control group, and should be considered supportive.

7.2. Glabellar Frown Lines (GFL)

The sponsor performed two placebo-controlled pivotal studies for the GFL indication (0724 and 0741), which are discussed below. Four supportive GFL studies were also submitted.

7.2.1. Pivotal Study 0724

7.2.1.1. Study design, objectives, locations and dates

This Phase III pivotal study was titled: 'A prospective, randomized, double-blind, placebo-controlled, multicentre trial to investigate the efficacy and safety of NT 201, free of complexing proteins, in the treatment of glabellar frown lines.'

It took place from 13 October 2008 to 15 June, 2009, in eight centres in the USA and Canada.

The primary objective was to assess the efficacy and safety of NT 201 in comparison to placebo.

7.2.1.2. Inclusion and exclusion criteria

Adults of either gender aged greater than or equal to 18 years were eligible if they had a diagnosis of moderate to severe glabellar frown lines (with a severity score of 2 or 3 on the Facial Wrinkle Scale, defined below). They also had to be medically stable and provide written informed consent.

Exclusion criteria included:

- previous botulinum toxin treatment in the last eight months;
- glabellar scars or facial surgery or implantation;
- fixed glabellar frown lines that could not be substantially lessened by spreading them apart;
- marked facial asymmetry or ptosis of eyelid or eyebrow;
- infection in the anticipated injection sites;
- any neuromuscular condition such as myasthenia or Lambert-Eaton syndrome
- previous Bell's palsy;
- use of drugs affecting neuromuscular transmission, such as aminoglycosides;
- bleeding disorders, or use of drugs affecting haemostasis, including anticoagulants and antiplatelets, within the last 14 days;
- use of aminoquinolones (chloroquine and related compounds) that might antagonise the onset of action of NT 201;
- recent alcohol or drug abuse;
- at risk of pregnancy;
- hypersensitivity to trial medication;
- family history of sudden death or prolonged QT syndrome;
- involvement in other experimental studies;
- potential conflicts of interest, such as employees of institutions involved in the study, or their relatives.

Note that previous treatment with botulinum toxin was permitted, provided that it occurred early enough before recruitment so as not to interfere with the assessment. About a third of subjects had received some prior botulinum toxin therapy.

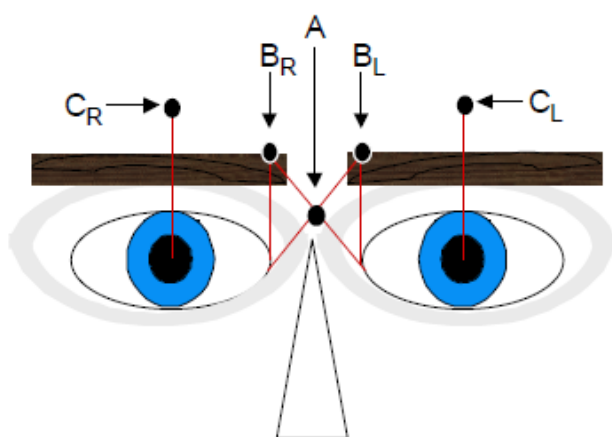
On the whole, these are standard and reasonable exclusion criteria, aimed at removing potential confounding factors and minimising safety risks.

7.2.1.3. Study treatments

Subjects received NT 201 or matching placebo, as a single intramuscular dose of 20 U (0.5 mL) on Day 0, reconstituted in 0.9% sodium chloride (NaCl) and distributed in equal portions to five injection sites,:

- procerus muscle;
- medial right and left corrugator muscle approximately 1 cm above the bony orbital rim;
- middle of right and left corrugator muscle at least 1.5 cm above the bony orbital rim.

Figure 5: Positioning of NT 201 injection sites



The placebo contained the excipients of NT 201, without any toxin.

Subjects received a single dose but were followed for 120 days.

7.2.1.4. Efficacy variables and outcomes

Each subject underwent seven visits: Screening Visit (Day -14 to -3), Visit 1 (Day 0), Visit 2 (Day 7 plus or minus 3), Visit 3 (Day 30 plus or minus 7), Visit 4 (Day 60 plus or minus 7), Visit 5 (Day 90 plus or minus 7), Visit 6 (Day 120 plus or minus 7). At each visit, the degree of facial/glabellar wrinkling was assessed as follows.

Physicians assessed patients with the Facial Wrinkle scale (FWS), which was a four-point categorical scale based on the investigator's assessment of the severity of glabellar frown lines at maximum frown: 'none' (0), 'mild' (1), 'moderate' (2), 'severe' (3).

Subjects also entered a subjective Patient's Assessment on 4-point scale, in answer to the following question about their maximum frown: 'How would you judge the potency of frown muscle action by comparison to sample photos at this visit?' Possible responses were:

- 0 = No muscle action at all;
- 1 = Some or slight muscle action possible that is visible furrows;
- 2 = Moderately strong muscle action possible that is visible muscle bulges;
- 3 = Strong muscle action possible which may cause local pallor.

'Two-point responders' for either variable were defined as subjects with a 2-point improvement on Day 30 compared to Day 0.

The primary efficacy variable was a composite endpoint (CETS, Composite endpoint treatment success), which was generated from the Facial Wrinkle scale (FWS) and the Patient's

Assessment on a 4-point scale, in which subjects were considered treatment successes if they were 2-point responders for both variables.

Secondary variables were listed by the sponsor as follows:

- Percentage of none/mild responders at rest at Day 30 by investigator's rating on FWS. (Responder was defined as a subject with a rating of none or mild).
- Percentage of 1-point responders at rest at Day 30 by Patient's Assessment on 4-point scale. (Responder was defined as a subject with 1-point improvement compared to baseline.)
- Percentage of none/mild responders at maximum frown at Day 30 by investigator's rating on FWS. (Responder was defined as a subject with the rating none or mild.)
- Percentage of 1-point responders at maximum frown at Day 30 by Patient's Assessment on 4-point scale. (Responder was defined as a subject with 1-point improvement compared to baseline).

Tertiary efficacy variables were the course over time of the percentages of responders by the above definitions, and the percentage of responders as assessed by an independent expert committee using digitised photographs of the subjects.

7.2.1.5. Randomisation and blinding methods

Subjects were randomised with a computerised randomisation program, RANCODE (Version 3.6, IDV Datenanalyse und Versuchsplanung, Gauting, Germany), by a randomisation officer, with randomisation codes kept sealed and separate from the treating institutions.

Randomisation was performed with a 2:1 ratio, such that subjects were twice as likely to receive active treatment.

7.2.1.6. Analysis populations

The sponsor identified two major study populations: the Full Analysis Set (FAS), which consisted of all subjects who were randomised and treated with study medication, and the Per Protocol Set (PPS), which consisted of all subjects in the FAS who had:

- investigator assessments for severity of glabellar frown lines at maximum frown at Day 30;
- the patient's subjective assessment (4-point scale) at Day 30; and
- no major protocol deviations.

For the safety evaluation, the sponsor defined a Safety Evaluation Set (SES), which was identical to the FAS.

7.2.1.7. Statistical methods

The primary efficacy variable (CETS) was based on the FAS (all subjects who were randomised and treated). A Cochran-Mantel-Haenszel (CMH) test was intended to be applied to the percentage of subjects successful in the CETS at Day 30 in the treatment group in comparison to the placebo group, using centre as a stratification variable. Superior efficacy over placebo was to be concluded if the two sided p-value was greater than or equal to 0.05. Because there were zero responders in the placebo group, it was not possible to perform the intended CMH procedure. Instead, comparison between the treatment groups was performed by the Fisher's exact test only, the p-value of this test was considered as confirmatory. Given the complete lack of efficacy in the placebo group by the prospectively defined primary outcome, this approach was reasonable, and indeed it seems unlikely that the interpretation of the outcome hinges on the exact statistical method used.

Imputation of missing values was performed with a 'last observation carried forward' (LOCF) approach. Given that NT 201 is expected to produce an early benefit that then wanes over subsequent weeks, an LOCF approach could potentially lead to imputation of maximal

responses to later times when the actual response had waned. This problem is offset by assessing the primary outcome at Day 30, near the expected time of peak effect. Also, the sponsor performed sensitivity analyses to determine the influence of various methods of imputing missing values.

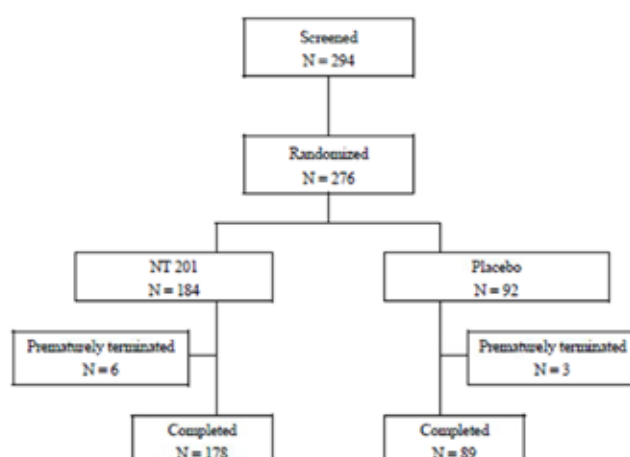
Secondary and tertiary efficacy variables were only presented with descriptive statistics, rather than comparative statistics, which is appropriate given their lesser importance.

7.2.1.8. Sample size

The sponsor calculated the sample size for a CMH test using 'nQuery Advisor', using a two sided alpha of 0.05, and an odds ratio OR to be detected of 4.5 (calculated with $P_{NT201} = 0.25$ and $P_{Placebo} = 0.07$), based on a randomisation ratio of 2:1 for NT 201:placebo. (Because the respective module in 'nQuery advisor' only calculates sample sizes for the CMH situation for equal sample sizes, a correction factor of 1.13 was applied; this correction factor was derived using the nQuery module chi square tests, under the assumption that similar power adjustments would be necessary for the different statistical tests. This assumption may not have been valid, but the point is academic because the placebo response rate was zero and the CMH test could not be applied after all.) Under these assumptions, a power of more than 90% was achieved for a study size of at least 255 subjects (that is, 170:85 to NT 201 and placebo respectively). Allowing for a screening failure rate of 10% about 285 subjects had to be screened; the final number screened exceeded this ($n = 294$) and the number of subjects randomised ($n = 276$) was above target.

7.2.1.9. Participant flow

Participant flow is displayed in Figure 6 and Table 5 below. Six NT 201 recipients and three placebo recipients terminated their involvement prematurely, due to withdrawal of consent or loss of follow-up. Thus, withdrawals were minimal and in proportion to the original 2:1 randomisation ratios, indicating very little chance of substantial withdrawal bias.

Figure 6: Disposition of subjects

Source: Table 14.1.1.1; Table 14.1.1.3.

Table 5: Disposition of subjects and reasons for discontinuations – SES/FAS

	NT 201 (N=184)		Placebo (N=92)		Total (N=276)	
	n	(%)	n	(%)	n	(%)
Completed study (Visit 6)	178	(96.7)	85	(96.7)	267	(96.7)
Completed Visit 3	180	(97.8)	91	(98.9)	271	(98.2)
Prematurely terminated	6	(3.3)	3	(3.3)	5	(3.3)
Reason for premature termination						
Lack of efficacy	0	(0.0)	0	(0.0)	0	(0.0)
Death	0	(0.0)	0	(0.0)	0	(0.0)
Adverse Event(s)	1	(0.5)	0	(0.0)	1	(0.4)
Consent withdrawn	3	(1.6)	3	(3.3)	6	(2.2)
Lost to follow-up	3	(1.6)	0	(0.0)	3	(1.1)
Non-compliance	0	(0.0)	0	(0.0)	0	(0.0)
Administrative reason(s)	0	(0.0)	0	(0.0)	0	(0.0)

FAS = Full Analysis Set; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N; SES = Safety Evaluation Set.

Note: More than one reason for premature termination possible.

In this study SES was identical to FAS.

Visit 3 = Day 30; Visit 6 = Day 120.

7.2.1.10. Major protocol violations/deviations

Protocol deviations are summarised below. The number of protocol violations was acceptable for a study of this nature and the percentage of subjects excluded from the PPS was similar in the two treatment groups.

Table 6: Protocol deviations – reasons for exclusion from per protocol set

	NT 201 (N=184)	Placebo (N=92)	Total (N=276)
	n (%)	n (%)	n (%)
Number of subjects excluded from the Per Protocol Set (PPS)	17 (9.2)	7 (7.6)	24 (8.7)
Reason for exclusion ¹			
Forbidden concomitant medication	1 (0.5)	1 (1.1)	2 (0.7)
Forbidden prior medication	2 (1.1)	0 (0.0)	2 (0.7)
Missing primary efficacy parameter	4 (2.2)	1 (1.1)	5 (1.8)
Missing secondary efficacy parameter	4 (2.2)	1 (1.1)	5 (1.8)
Other (Positive pregnancy test)	1 (0.5)	0 (0.0)	1 (0.4)
Rater change	3 (1.6)	1 (1.1)	4 (1.4)
Violation of visit window	7 (3.8)	4 (4.3)	11 (4.0)

N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N.

¹ Excluded subjects might have more than one reason for exclusion.**7.2.1.11. Baseline data**

Baseline characteristics are summarised in the table below. There were no substantial baseline differences between the active and placebo groups.

Table 7: Demographic data and other baseline characteristics (SES/FAS)

	NT 201 (N=184)	Placebo (N=92)	Total (N=276)
Age categories (n (%))			
≤50 years	121 (65.8)	55 (59.8)	176 (63.8)
>50 years	63 (34.2)	37 (40.2)	100 (36.2)
Sex (n (%))			
Male	22 (12.0)	16 (17.4)	38 (13.8)
Female	162 (88.0)	76 (82.6)	238 (86.2)
Ethnic Origin (n (%))			
White	163 (88.6)	83 (90.2)	246 (89.1)
Black or African American	3 (1.6)	3 (3.3)	6 (2.2)
Asian	10 (5.4)	1 (1.1)	11 (4.0)
Hispanic or Latino	7 (3.8)	3 (3.3)	10 (3.6)
American Indian or Alaska Native	0 (0.0)	1 (1.1)	1 (0.4)
Native Hawaiian, other Pacific Islander	1 (0.5)	0 (0.0)	1 (0.4)
Other	0 (0.0)	1 (1.1)	1 (0.4)
Pre-treatment with botulinum neurotoxin¹			
Yes	60 (32.6)	25 (27.2)	85 (30.8)
No	124 (67.4)	67 (72.8)	191 (69.2)
Age [years]			
n	184	92	276
Mean (SD)	46.6 (9.87)	46.4 (10.56)	46.6 (10.09)
Median	46.0	48.0	47.0
Min, Max	25, 72	20, 67	20, 72
Height [cm]			
n	184	92	276
Mean (SD)	166.7 (8.74)	167.5 (8.82)	166.9 (8.76)
Median	165.0	167.0	165.5
Min, Max	149, 194	150, 193	149, 194
Weight [kg]			
n	184	92	276
Mean (SD)	69.09 (13.882)	70.09 (13.958)	69.42 (13.890)
Median	68.00	68.50	68.00
Min, Max	45.0, 121.0	47.0, 113.0	45.0, 121.0
BMI [kg/m²]			
n	184	92	276
Mean (SD)	24.80 (4.258)	24.88 (3.907)	24.83 (4.137)
Median	24.00	24.20	24.00
Min, Max	16.9, 41.5	18.7, 34.9	16.9, 41.5

FAS = Full Analysis Set, Max = Maximum, Min = Minimum, N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N; SD = Standard deviation; SES = Safety Evaluation Set.

Note: In this study SES is identical to FAS.

¹ As allowed according to protocol.

7.2.1.12. Results for the primary efficacy outcome

Results for the primary endpoint (CETS) are shown in the table below. In the NT 201 group, 60.3% of subjects were CETS responders at Day 30, whereas no subject in the placebo group was a CETS responder. In fact, no placebo recipient showed a two-point response for either component of the CETS. The lack of response in the placebo group made it impossible to perform the CMH procedure, so the sponsor performed a post hoc analysis with the Fisher's exact test. This reached a high degree of statistical significance (p-value < 0.0001). Changing statistical tests post hoc is normally considered inappropriate, but the sponsor's approach was reasonable under these circumstances, and the differences between the groups are so marked that it is unlikely that the choice of statistical test made a substantial difference. The 95% CI for the difference in response rate (RR) was 0.52 to 0.68.

For the individual efficacy variables, two-point response rates were slightly better than the combined response rate (76.6% for FWS, 65.2% for Patient Assessment), and the differences were statistically significant (p-value < 0.0001; Fisher's exact test).

Table 8: Study 0724 primary efficacy analysis – CETS: responded analysis – imputed data (FAS)

	20 U NT 201 N=184	Placebo N=92
CETS		
Responder [n (%)]	111 (60.3)	0 (0.0)
Response rate imputed (RR)		0.60
Difference RR (NT 201-Placebo)		0.60
95% CI for difference RR		0.52, 0.68
p-value (Fisher's exact test)		<0.0001
2-point response FWS		
Responder [n (%)]	141 (76.6)	0 (0.0)
Response rate imputed (RR)		0.77
Difference RR (NT 201-Placebo)		0.77
95% CI for difference RR		0.70, 0.84
p-value (Fisher's exact test)		<0.0001
2-point response patient assessment		
Responder [n (%)]	120 (65.2)	0 (0.0)
Response rate imputed (RR)		0.65
Difference RR (NT 201-Placebo)		0.65
95% CI for difference RR		0.58, 0.73
p-value (Fisher's exact test)		<0.0001

CETS: Composite endpoint treatment success; CI: Confidence interval; FAS: Full Analysis Set; FWS: Facial wrinkle scale; N: Number of subjects in specific group; n: Number of subjects; RR: Response rate; Calculation of percentages based on N.

If the analysis is limited to observed FAS cases (leaving out cases where imputation was necessary), or observed PPS cases (leaving out cases with protocol deviations), broadly similar results are obtained, as shown in the tables below.

Table 9: CETS: responded analysis at Day 30 – observed cases - FAS

	NT 201 (N=184)	Placebo (N= 92)
CETS		
Response: Yes [n (%)]	108 (58.7)	0 (0.0)
Response: No [n (%)]	72 (39.1)	91 (98.9)
Missing [n (%)]	4 (2.2)	1 (1.1)
Response rate (RR)	0.60	0.00
Difference RR (NT 201-Placebo)	0.60	
95% CI for difference RR	0.52, 0.68	
p-value (Fisher's exact test; descriptive)	<0.0001	
2-point response FWS		
Response: Yes [n (%)]	138 (75.0)	0 (0.0)
Response: No [n (%)]	42 (22.8)	91 (98.9)
Missing [n (%)]	4 (2.2)	1 (1.1)
Response rate (RR)	0.77	0.00
Difference RR (NT 201-Placebo)	0.77	
95% CI for difference RR	0.70, 0.84	
p-value (Fisher's exact test; descriptive)	<0.0001	
2-point response patient's assessment		
Response: Yes [n (%)]	117 (63.6)	0 (0.0)
Response: No [n (%)]	63 (34.2)	91 (98.9)
Missing [n (%)]	4 (2.2)	1 (1.1)
Response rate (RR)	0.65	0.00
Difference RR (NT 201-Placebo)	0.65	
95% CI for difference RR	0.57, 0.73	
p-value (Fisher's exact test; descriptive)	<0.0001	

CETS = Composite endpoint treatment success; CI = Confidence interval; FAS = Full Analysis Set; FWS = Facial wrinkle scale; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N.

Table 10: CETS: Responder analysis at Day 30 – observed cases - PPS

	NT 201 (N=167)	Placebo (N=85)
CETS		
Response: Yes [n (%)]	102 (61.1)	0 (0.0)
Response: No [n (%)]	65 (38.9)	85 (100.0)
Response rate (RR)	0.61	0.00
Difference RR (NT 201-Placebo)	0.61	
95% CI for difference RR	0.53, 0.69	
p-value (Fisher's exact test; descriptive)	<0.0001	
2-point response FWS		
Response: Yes [n (%)]	129 (77.2)	0 (0.0)
Response: No [n (%)]	38 (22.8)	85 (100.0)
Response rate (RR)	0.77	0.00
Difference RR (NT 201-Placebo)	0.77	
95% CI for difference RR	0.70, 0.84	
p-value (Fisher's exact test; descriptive)	<0.0001	
2-point response patient's assessment		
Response: Yes [n (%)]	110 (65.9)	0 (0.0)
Response: No [n (%)]	57 (34.1)	85 (100.0)
Response rate (RR)	0.66	0.00
Difference RR (NT 201-Placebo)	0.66	
95% CI for difference RR	0.58, 0.74	
p-value (Fisher's exact test; descriptive)	<0.0001	

CETS = Composite endpoint treatment success; CI = Confidence interval; FWS = Facial wrinkle scale; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N; PPS = Per Protocol Set.

7.2.1.13. Results for other efficacy outcomes

The sponsor also assessed none/mild response rates, defined as those achieving an investigator FWS rating of 'none' or 'mild' at Day 30, regardless of baseline rating. At maximum frown,

almost 80% of NT 201 recipients and no placebo recipients were responders by this definition, a significant difference by Fisher's exact test.

Table 11: Percentage of responders at maximum frown at Day 30 (investigator's rating on FWS) – observed cases - FAS

	NT 201 (N=184)	Placebo (N=92)
Response		
Yes [n (%)]	147 (79.9)	0 (0.0)
No [n (%)]	33 (17.9)	91 (98.9)
Missing [n (%)]	4 (2.2)	1 (1.1)
Response rate (RR)	0.82	0.00
Difference RR (NT 201-Placebo)	0.82	
95% CI for difference RR	0.75, 0.88	
Odds ratio (OR)	NA	
Exact 95% CI for OR	NA	
p-value (Fisher's exact test; descriptive)	<0.0001	

CI = Confidence interval; FAS = Full Analysis Set; FWS = Facial Wrinkle Scale; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N; NA= Not applicable.

Note: Visit 3 = Day 30.

A subject is considered as responder at maximum frown, if the investigator's rating on the FWS is none or mild.

Response rates are calculated based on the number of subjects with available response information.

For the Patient's Assessment, the primary analysis was based on two-point responses, but the sponsor also assessed one-point improvements at maximum frown. Response rates by this looser definition were seen more commonly than with more restrictive definitions, but there was still a marked difference between treatment groups.

Table 12: Percentage of one 1-point responders at maximum frown at Day 30 (patient's assessment on 4-point scale) – observed cases – FAS

	NT 201 (N=184)	Placebo (N=92)
1-point response		
Yes [n (%)]	161 (87.5)	9 (9.8)
No [n (%)]	19 (10.3)	82 (89.1)
Missing [n (%)]	4 (2.2)	1 (1.1)
Response rate (RR)	0.89	0.10
Difference RR (NT 201-Placebo)	0.80	
95% CI for difference RR	0.71, 0.88	
Odds ratio (OR)	77.20	
Exact 95% CI for OR	31.53, 198.32	
p-value (Fisher's exact test; descriptive)	<0.0001	

FAS = Full Analysis Set; CI = Confidence interval; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N.

Response rate at rest showed a lesser difference between treatment groups, as expected for a treatment that works by causing muscle weakness (see tables below). These endpoints included a substantial placebo response, but nonetheless showed a statistically significant treatment benefit according to Fisher's exact test. (The statistical analysis of these minor endpoints was considered descriptive rather than confirmatory).

Table 13: Percentage of 1-point responders at rest at Day 30 (patient's assessment on 4-point scale) observed cases - FAS

	NT 201 (N=184)	Placebo (N=92)
1-point response		
Yes [n (%)]	131 (71.2)	10 (10.9)
No [n (%)]	49 (26.6)	81 (88.0)
Missing [n (%)]	4 (2.2)	1 (1.1)
Response rate (RR)	0.73	0.11
Difference RR (NT 201-Placebo)	0.62	
95% CI for difference RR	0.52, 0.72	
Odds ratio (OR)	21.66	
Exact 95% CI for OR	10.03, 50.06	
p-value (Fisher's exact test; descriptive)	<0.0001	

CI = Confidence interval; FAS = Full Analysis Set; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N.

Responses at times other than Day 30 were considered to be tertiary, but give a useful insight into the temporal profile of the response. Using the two-point response rate for FWS, NT 201 administered for GFL appears to produce its maximal effect in the first week, show persistence of efficacy for the first month, and then begin to wane in efficacy by the second and third month

(see figure below). Using an alternative definition of response (subjects achieving an FWS rating of 'one' or 'mild') produced a very similar response profile (subsequent figure). By either definition, the placebo group showed no response at any time. Assessments at rest (one-point improvement in FWS) were associated with a moderate placebo effect, but the overall time profile in the active group (third figure below) was similar to that seen with more restrictive definitions. Two-point responses on the subjective patient-assessment scale showed a similar profile to investigator-based responses (fourth figure below). Restricting the analysis to observed cases or to the PPS population did not make any substantial difference (not shown).

Figure 7: Course over time of percentages of 2-point responders at maximum from by investigator's rating on FWS at Days 7, 30, 60, 90, 120 – observed cases - FAS

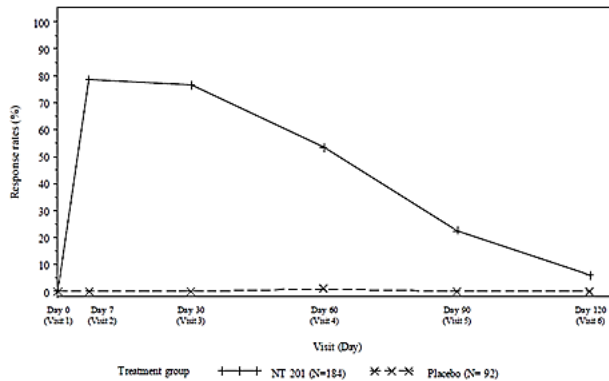


Figure 8: Course over time of percentages of responders at maximum from by investigator's rating on FWS at Days 7, 30, 60, 90, 1210 – observed cases - FAS

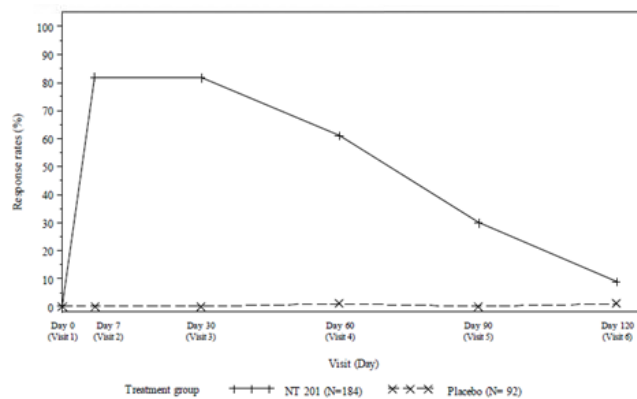


Figure 9: Course over time of percentages of responders at rest by investigator's rating on FWS at Days 7, 30, 60, 90, 1210 – observed cases - FAS

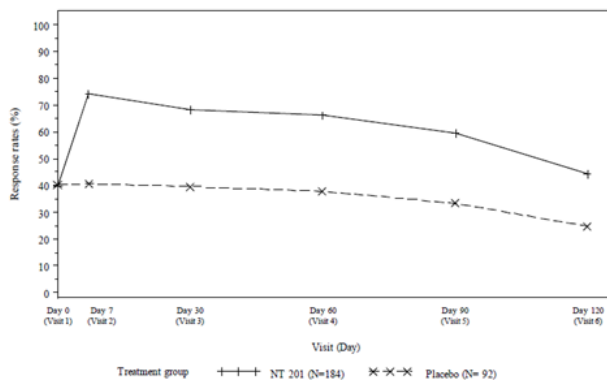
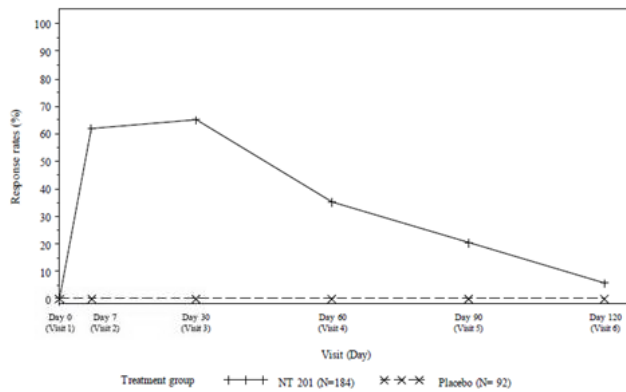


Figure 10: Course over time of percentages of responders at maximum from by patient's assessment on 4-point scale at Days 7, 30, 60, 90, 1210 – observed cases - FAS



Note that, in this analysis, the response rate in the few days prior to the first post-treatment visit (Visit 2 at Day 7) is unknown. Some information about the treatment response in the first week was captured by asking patients to note the subjective onset of a treatment effect. In the active group, the median day of onset of effect was Day 3, which represents two days elapsed since the injection. In the placebo group, very few subjects reported an onset of any effect. The cumulative proportion of subjects reporting an onset of treatment effect is shown in the subsequent figure.

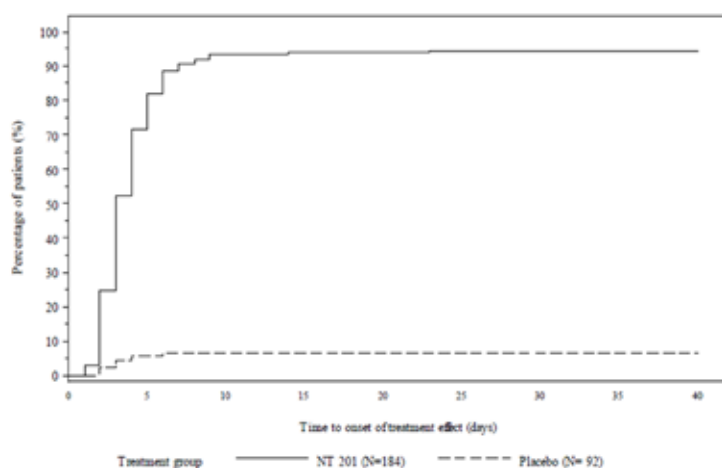
Table 14: Time to onset of treatment effect – observed cases - FAS

	NT 201 (N=184)	Placebo (N=92)
Onset of treatment effect (recognized by subject until Visit 3)		
Yes n (%)	173 (94.0)	6 (6.5)
No n (%)	10 (5.4)	85 (92.4)
Missing n (%)	1 (0.5)	1 (1.1)
Time to onset of treatment effect (days)		
n	173	6
Nmiss	11	86
Mean	3.8	3.3
SD	2.31	1.51
Min	1	2
Q1	2.0	2.0
Median	3.0	3.0
Q3	4.0	4.0
Max	23	6
Time to onset of treatment effect		
1 day n (%)	5 (2.7)	0 (0.0)
2 days n (%)	40 (21.7)	2 (2.2)
3 days n (%)	51 (27.7)	2 (2.2)
4 days n (%)	35 (19.0)	1 (1.1)
5 days n (%)	19 (10.3)	0 (0.0)
6 days n (%)	12 (6.5)	1 (1.1)
7 days n (%)	4 (2.2)	0 (0.0)
8 days n (%)	2 (1.1)	0 (0.0)
9 days n (%)	3 (1.6)	0 (0.0)
>10 days n (%)	2 (1.1)	0 (0.0)
Missing n (%)	11 (6.0)	86 (93.5)

FAS = Full Analysis Set; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N; SD = Standard deviation.

Note: Visit 3 = Day 30.

Time to onset of treatment effect = Date of onset of treatment effect - date of study medication + 1. The time to onset equals 1 if the treatment effect was noticed at the day of injection. This means a median onset of 3.0 days is actually an onset 2 days post injection.

Figure 11: Time to onset of treatment effect – observed cases - FAS

Finally, in an assessment of the reproducibility and robustness of the results, digital photos of the subjects were assessed by an independent expert committee. The two-point response rates from the committee assessments (74.5% for observed cases in the active group, 0% in the placebo group) were similar to the response rates derived from investigator assessments and showed a significant treatment effect ($p < 0.0001$ by Fisher's exact test, see Table 15 below). Using the committee's ratings with an alternative definitions of response (1-point improvement or none/mild response) produced similar results to the investigator-based ratings (not shown).

Table 15: Percentage of 2-point responders at maximum frown at Day 30 by an independent expert committee according to FWS based on standardised digital photographs – observed cases - FAS

	NT 201 (N=184)	Placebo (N= 92)
2-point response		
Yes [n (%)]	137 (74.5)	0 (0.0)
No [n (%)]	42 (22.8)	88 (95.7)
Missing [n (%)]	5 (2.7)	4 (4.3)
Response rate (RR)	0.77	0.00
Difference RR (NT 201-Placebo)	0.77	
95% CI for difference RR	0.69, 0.84	
Odds ratio (OR)	NA	
Exact 95% CI for OR	NA	
p-value (Fisher's exact test; descriptive)	<0.0001	

CI = Confidence interval; FAS = Full Analysis Set; FWS = Facial Wrinkle Scale; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N; NA= Not applicable.

Note: Visit 3 = Day 30.

Response rates are calculated based on the number of subjects with available response information. A 2-point responder at maximum frown is a subject with an improvement of at least 2 points on the FWS compared to baseline. The final variable containing the response Yes/No for each subject, visit and situation was generated by applying majority vote on the three derived auxiliary response variables from the experts. That means, the subject (with respective visit and situation) was considered as 2-responder in case the subject had an improvement of at least 2 points in three or two expert ratings.

7.2.1.14. Subgroup analyses

The sponsor performed subgroup analyses by gender, age, ethnicity and pre-treatment status, which showed a significant treatment effect in all subgroups. This is not surprising given the lack of any responders in the placebo group for the primary efficacy outcome. The subgroup analysis was repeated in Study 041, discussed below, and also in the pooled analysis of both studies. Because the results were consistent across the two studies, the overall subgroup analysis was considered. The sponsor also assessed response according to baseline disease severity. A significant response was seen in all subgroups except in the small, underpowered group rated by the investigator as moderate (active n = 23, placebo n = 6); even in this group the trend was favourable (active response rate 43.5%, placebo response rate 0%, p = 0.0676, Fisher's exact test).

7.2.2. Pivotal Study 0741

7.2.2.1. Study design, objectives, locations and dates

This study bore the same title as Study 0724: 'A prospective, randomised, double-blind, placebo-controlled, multicenter trial to investigate the efficacy and safety of NT 201, free of complexing proteins, in the treatment of glabellar frown lines.' The design and objectives of the two studies were identical. It was performed in eight centres in the USA, from 10 October 2008 to 19 June 2009.

7.2.2.2. Inclusion and exclusion criteria

Inclusion and exclusion criteria were as described for Study 0724. The study population consisted of adults of either gender with moderate or severe GFL.

7.2.2.3. Study treatments

Subjects received 20 U NT 201 or matching placebo, injected as a single dose to five sites in the brow region, as described for Study 0724. Subjects received a single dose but were followed for 120 days.

7.2.2.4. Efficacy variables and outcomes

The primary endpoint was the treatment success rate (CETS) as described for Study 0724; this was based on the FWS and Patient Assessment. More inclusive definitions of response were also considered as secondary and tertiary endpoints, as in Study 0724. An independent committee applied the FWS to digital photos taken on Day 30; these scores were considered a tertiary endpoint.

7.2.2.5. *Randomisation and blinding methods*

Subjects were randomised in a 2:1 ratio to active treatment or placebo, as described for Study 0724. Blinding was maintained by using identical vials and excipients for the active and placebo preparations, and by keeping randomisation codes in a sealed and secure location.

7.2.2.6. *Analysis populations*

As in Study 0724, the sponsor defined two populations for the efficacy analysis, the FAS and PPS. For the safety analysis, the FAS population was used, but in this context was called the SES.

7.2.2.7. *Sample size*

Sample size calculations were as described for Study 0724.

7.2.2.8. *Statistical methods*

The sponsor intended to use the CMH procedure with a 0.05 significance level, but was unable to apply this test because the response rate in the placebo group was zero. Instead, Fisher's exact test was used. Under the circumstances, this is reasonable.

The primary analysis was performed on the FAS, but supportive analyses were performed with the PPS and produced similar conclusions. Data imputation was performed with a LOCF approach, but additional analyses based on observed cases were also performed, and produced similar results.

Overall, the statistical methods were appropriate, and the magnitude of the treatment effect was large enough that it is very unlikely that interpretation of the results depends substantially on the statistical methods chosen.

7.2.2.9. *Participant flow*

Disposition of subjects is displayed in the figure and table below. Overall, a high proportion of subjects completed the study, and withdrawals are unlikely to have had a significant effect on the outcome.

Figure 12: Disposition of subjects

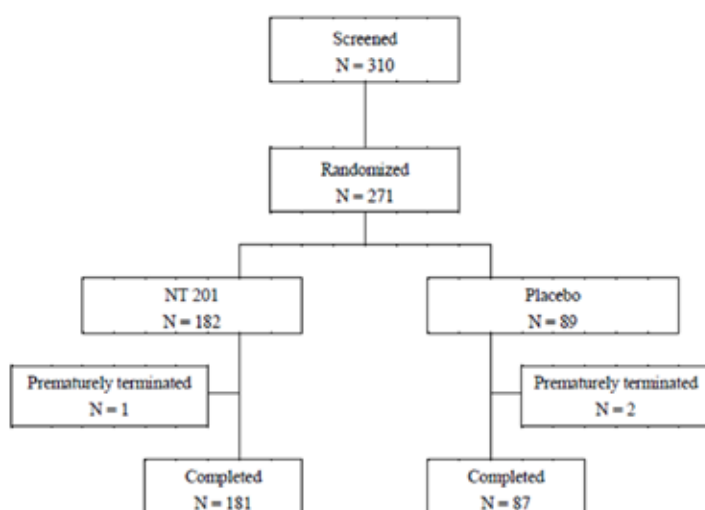


Table 16: Disposition of subjects and reasons for discontinuations – SES/FAS

	NT 201 (N=182)		Placebo (N=89)		Total (N=271)	
	n	(%)	n	(%)	n	(%)
Completed study (Visit 6)	181	(99.5)	87	(97.8)	268	(98.9)
Completed Visit 3	182	(100.0)	87	(97.8)	269	(99.3)
Prematurely terminated	1	(0.5)	2	(2.2)	3	(1.1)
Reason for premature termination						
Lack of efficacy	0	(0.0)	0	(0.0)	0	(0.0)
Death	0	(0.0)	0	(0.0)	0	(0.0)
Adverse Event(s)	0	(0.0)	0	(0.0)	0	(0.0)
Consent withdrawn	1	(0.5)	0	(0.0)	1	(0.4)
Lost to follow-up	0	(0.0)	2	(2.2)	2	(0.7)
Non-compliance	0	(0.0)	0	(0.0)	0	(0.0)
Administrative reason(s)	0	(0.0)	0	(0.0)	0	(0.0)

FAS = Full Analysis Set; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N; SES = Safety Evaluation Set.

7.2.2.10. Major protocol violations/deviations

Protocol deviations are summarised in the table below. Overall, the number of violations was low, and acceptable for a study of this nature

Table 17: Protocol deviations – reasons for exclusion from per protocol set

	NT 201 (N=182)		Placebo (N=89)		Total (N=271)	
	n	(%)	n	(%)	n	(%)
Number of subjects excluded from the Per Protocol Set (PPS)	12	(6.6)	9	(10.1)	21	(7.7)
Reason for exclusion ¹						
EXCL14: Any regular intake of drugs in doses leading to anticoagulative effect	0	(0.0)	1	(1.1)	1	(0.4)
EXCL18: Concurrent use of aminoglycoside antibiotics or other agents that might interfere with neuromuscular function or with the action of botulinum toxin	1	(0.5)	0	(0.0)	1	(0.4)
Forbidden concomitant medication	6	(3.3)	3	(3.4)	9	(3.3)
Forbidden prior medication	1	(0.5)	1	(1.1)	2	(0.7)
Forbidden treatment	1	(0.5)	0	(0.0)	1	(0.4)
Missing primary efficacy parameter	0	(0.0)	2	(2.2)	2	(0.7)
Missing secondary efficacy parameter	0	(0.0)	2	(2.2)	2	(0.7)
Non-compliance with randomization procedure	1	(0.5)	0	(0.0)	1	(0.4)
Rater change	0	(0.0)	1	(1.1)	1	(0.4)
Violation of visit window	4	(2.2)	2	(2.2)	6	(2.2)

N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N.

¹ Excluded subjects might have more than one reason for exclusion.

7.2.2.11. Baseline data

Baseline characteristics are summarised in the table below. There were no substantial baseline differences between the active and placebo groups.

Table 18: Demographic data and other baseline characteristics (SES/FAS)

	NT 201 (N=182)	Placebo (N=89)	Total (N=271)
Age categories (n (%))			
≤50 years	119 (65.4)	59 (66.3)	178 (65.7)
>50 years	63 (34.6)	30 (33.7)	93 (34.3)
Sex (n (%))			
Male	12 (6.6)	5 (5.6)	17 (6.3)
Female	170 (93.4)	84 (94.4)	254 (93.7)
Ethnic origin (n (%))			
White	119 (65.4)	58 (65.2)	177 (65.3)
Black or African American	3 (4.4)	3 (3.4)	11 (4.1)
Asian	2 (1.1)	0 (0.0)	2 (0.7)
Hispanic or Latino	51 (28.0)	28 (31.5)	79 (29.2)
American Indian or Alaska Native	1 (0.5)	0 (0.0)	1 (0.4)
Other	1 (0.5)	0 (0.0)	1 (0.4)
Pre-treatment with botulinum neurotoxin¹			
Yes	39 (21.4)	13 (14.6)	52 (19.2)
No	143 (78.6)	76 (85.4)	219 (80.8)
Age [years]			
n	182	89	271
Mean (SD)	46.9 (9.32)	45.7 (11.41)	46.5 (10.05)
Median	46.5	46.0	46.0
Min, Max	24, 74	18, 71	18, 74
Height [cm]			
n	182	89	271
Mean (SD)	164.5 (7.90)	163.5 (6.98)	164.2 (7.61)
Median	163.0	163.0	163.0
Min, Max	149, 198	147, 198	147, 198
Weight [kg]			
n	182	89	271
Mean (SD)	68.67 (13.832)	71.55 (14.809)	69.62 (14.198)
Median	65.00	68.00	66.00
Min, Max	45.0, 118.0	49.0, 118.0	45.0, 118.0
BMI [kg/m²]			
n	182	89	271
Mean (SD)	25.36 (4.601)	26.88 (6.069)	25.86 (5.169)
Median	24.55	25.70	24.80
Min, Max	16.3, 44.4	18.3, 49.1	16.3, 49.1

FAS = Full Analysis Set; Max = Maximum; Min = Minimum; N = Number of subjects in specific group;
n = Number of subjects; Calculation of percentages based on N; SD = Standard deviation; SES = Safety Evaluation Set.

7.2.2.12. Results for the primary efficacy outcome

The treatment success rate (CETS) in this study was lower than in the similarly designed Study 0724 – 48% rather than 60%, but in both studies the placebo success rate was zero. The investigator-based FWS showed that most subjects (70.9%) were two-point responders, but the Patient's Assessment was less favourable, with only 55.5% showing a two-point response. All comparisons showed clear statistical superiority of NT 201 over placebo by Fisher's exact test ($p < 0.0001$).

Table 19: CETS: Responder analysis at Day 30 – imputation of missing data - FAS

	NT 201 (N=182)	Placebo (N=89)
CETS		
Response: Yes [n (%)]	87 (47.8)	0 (0.0)
Response: No [n (%)]	95 (52.2)	89 (100.0)
Response rate imputed (RR)	0.48	0.00
Difference RR (NT 201-Placebo)	0.48	
95% CI for difference RR	0.40, 0.56	
p-value (Fisher's exact test)	<0.0001	
2-point response FWS		
Response: Yes [n (%)]	129 (70.9)	0 (0.0)
Response: No [n (%)]	53 (29.1)	89 (100.0)
Response rate imputed (RR)	0.71	0.00
Difference RR (NT 201-Placebo)	0.71	
95% CI for difference RR	0.63, 0.78	
p-value (Fisher's exact test)	<0.0001	
2-point response patient assessment		
Response: Yes [n (%)]	101 (55.5)	1 (1.1)
Response: No [n (%)]	81 (44.5)	88 (98.9)
Response rate imputed (RR)	0.55	0.01
Difference RR (NT 201-Placebo)	0.54	
95% CI for difference RR	0.46, 0.63	
p-value (Fisher's exact test)	<0.0001	

CETS = Composite endpoint treatment success; CI = Confidence interval; FAS = Full Analysis Set; FWS = Facial wrinkle scale; N= Number of subjects in specific group; n= Number of subjects; Calculation of percentages based on N.

† If the analysis is repeated with observed FAS cases, or with observed FWS cases, the results are similar, as shown in the tables below.†

Table 20: CETS: Responder analysis at Day 30 – observed cases - FAS

	NT 201 (N=182)	Placebo (N=89)
CETS		
Response: Yes [n (%)]	87 (47.8)	0 (0.0)
Response: No [n (%)]	95 (52.2)	87 (97.8)
Missing [n (%)]	0 (0.0)	2 (2.2)
Response rate (RR)	0.48	0.00
Difference RR (NT 201-Placebo)	0.48	
95% CI for difference RR	0.40, 0.56	
p-value (Fisher's exact test; descriptive)	<0.0001	
2-point response FWS		
Response: Yes [n (%)]	129 (70.9)	0 (0.0)
Response: No [n (%)]	53 (29.1)	87 (97.8)
Missing [n (%)]	0 (0.0)	2 (2.2)
Response rate (RR)	0.71	0.00
Difference RR (NT 201-Placebo)	0.71	
95% CI for difference RR	0.63, 0.78	
p-value (Fisher's exact test; descriptive)	<0.0001	
2-point response patient's assessment		
Response: Yes [n (%)]	101 (55.5)	1 (1.1)
Response: No [n (%)]	81 (44.5)	86 (96.6)
Missing [n (%)]	0 (0.0)	2 (2.2)
Response rate (RR)	0.55	0.01
Difference RR (NT 201-Placebo)	0.54	
95% CI for difference RR	0.46, 0.63	
p-value (Fisher's exact test; descriptive)	<0.0001	

CETS = Composite endpoint treatment success; CI = Confidence interval; FAS = Full Analysis Set; FWS = Facial wrinkle scale; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N.

Table 21: CETS: Responder analysis at Day 30 – observed cases - PPS

	NT 201 (N=170)	Placebo (N=80)
CETS		
Response: Yes [n (%)]	80 (47.1)	0 (0.0)
Response: No [n (%)]	90 (52.9)	80 (100.0)
Response rate (RR)	0.47	0.00
Difference RR (NT 201-Placebo)	0.47	
95% CI for difference RR	0.39, 0.55	
p-value (Fisher's exact test; descriptive)	<0.0001	
2-point response FWS		
Response: Yes [n (%)]	119 (70.0)	0 (0.0)
Response: No [n (%)]	51 (30.0)	80 (100.0)
Response rate (RR)	0.70	0.00
Difference RR (NT 201-Placebo)	0.70	
95% CI for difference RR	0.62, 0.78	
p-value (Fisher's exact test; descriptive)	<0.0001	
2-point response patient's assessment		
Response: Yes [n (%)]	94 (55.3)	1 (1.3)
Response: No [n (%)]	76 (44.7)	79 (98.8)
Response rate (RR)	0.55	0.01
Difference RR (NT 201-Placebo)	0.54	
95% CI for difference RR	0.45, 0.63	
p-value (Fisher's exact test; descriptive)	<0.0001	

CETS = Composite endpoint treatment success; CI = Confidence interval; FWS = Facial wrinkle scale; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N; PPS = Per Protocol Set.

7.2.2.13. Results for other efficacy outcomes

As in Study 0724, additional endpoints consisted of response rates with looser definitions of response and response rates for facial wrinkling assessed at rest, instead of at maximum frown. The assessments at rest showed a significant treatment effect, but the difference between active and placebo treatments was not as marked as the difference at maximum frown. This is consistent with the action of NT 201 on muscle contraction, and the milder frown lines present at rest in most subjects.

Table 22: Percentage of responders at maximum frown at Day 30 (investigator's rating on FWS) – observed cases - FAS

	NT 201 (N=182)	Placebo (N=89)
Response		
Yes [n (%)]	139 (76.4)	0 (0.0)
No [n (%)]	43 (23.6)	87 (97.8)
Missing [n (%)]	0 (0.0)	2 (2.2)
Response rate (RR)	0.76	0.00
Difference RR (NT 201-Placebo)	0.76	
95% CI for difference RR	0.69, 0.83	
Odds ratio (OR)	NA	
Exact 95% CI for OR	NA	
p-value (Fisher's exact test; descriptive)	<0.0001	

CI = Confidence interval; FAS = Full Analysis Set; FWS = Facial Wrinkle Scale; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N; NA= Not applicable.

Note: Visit 3 = Day 30.

A subject is considered as responder at maximum frown, if the investigator's rating on the FWS is none or mild. Response rates are calculated based on the number of subjects with available response information.

Table 23: Percentage of 1-point responders at maximum frown at Day 30 (patient's assessment on 4-point scale) – observed cases - FAS

	NT 201 (N=182)	Placebo (N=89)
1-point response		
Yes [n (%)]	152 (83.5)	10 (11.2)
No [n (%)]	30 (16.5)	77 (86.5)
Missing [n (%)]	0 (0.0)	2 (2.2)
Response rate (RR)	0.84	0.11
Difference RR (NT 201-Placebo)	0.72	
95% CI for difference RR	0.63, 0.81	
Odds ratio (OR)	39.01	
Exact 95% CI for OR	17.31, 92.70	
p-value (Fisher's exact test; descriptive)	<0.0001	

FAS = Full Analysis Set; CI = Confidence interval; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N.

Note: Visit 3 = Day 30.

A subject is considered as 1-point-responder at rest, if patient's assessment on 4-point scale at maximum frown is 1-point improved compared to baseline. Response rates are calculated based on the number of subjects with available response information.

Table 24: Percentage of responders at rest at Day 30 (investigator's rating on FWS) – observed cases - FAS

	NT 201 (N=182)	Placebo (N=89)
Response		
Yes [n (%)]	157 (86.3)	50 (56.2)
No [n (%)]	25 (13.7)	37 (41.6)
Missing [n (%)]	0 (0.0)	2 (2.2)
Response rate (RR)	0.86	0.57
Difference RR (NT 201-Placebo)	0.29	
95% CI for difference RR	0.16, 0.41	
Odds ratio (OR)	4.65	
Exact 95% CI for OR	2.45, 8.85	
p-value (Fisher's exact test; descriptive)	<0.0001	

CI = Confidence interval; FAS = Full Analysis Set; FWS = Facial Wrinkle Scale; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N.

Note: Visit 3 = Day 30.

A subject is considered as responder at rest, if the investigator's rating on the FWS is none or mild. Response rates are calculated based on the number of subjects with available response information.

Table 25: Percentage of 1-point responders at rest at Day 30 (patient's assessment on 4-point scale) – observed cases – FAS

	NT 201 (N=182)	Placebo (N=89)
1-point response		
Yes [n (%)]	137 (75.3)	13 (14.6)
No [n (%)]	45 (24.7)	74 (83.1)
Missing [n (%)]	0 (0.0)	2 (2.2)
Response rate (RR)	0.75	0.15
Difference RR (NT 201-Placebo)	0.60	
95% CI for difference RR	0.50, 0.71	
Odds ratio (OR)	17.33	
Exact 95% CI for OR	8.46, 36.92	
p-value (Fisher's exact test; descriptive)	<0.0001	

CI = Confidence interval; FAS = Full Analysis Set; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N.

Note: Visit 3 = Day 30.

A subject is considered as 1-point-responder at rest, if patient's assessment on 4-point scale at rest is 1 point improved compared to baseline. Response rates are calculated based on the number of subjects with available response information.

The temporal profile of the response was broadly similar in this study as in Study 0724, but the response rates at Day 30 were slightly better than the response rates at Day 7, whereas in Study 0724 the earlier responses were slightly better. Similar profiles were obtained when plotting two-point FWS responses, 'none/mild' FWS responses and two-point patient responses at maximum frown. Responses at rest were associated with a substantial placebo effect, but the shape of the response curve in the active group was similar to that obtained at maximum frown (see figures below).

Figure 13: Course over time of percentages of 2-point responders at maximum frown by investigator's rating on FWS at Days 7, 30, 60, 90, 120 – observed cases - FAS

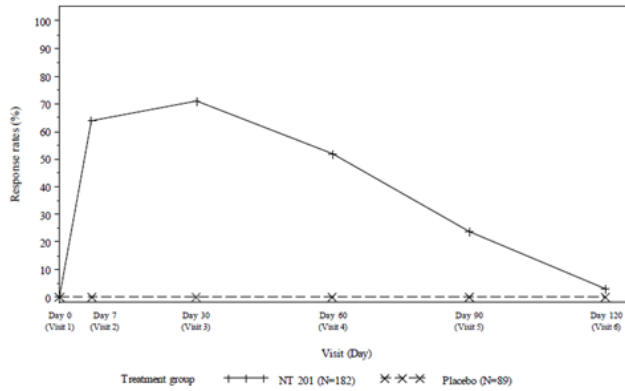


Figure 14: Course over time of percentages of responders at maximum frown by investigator's rating on FWS at Days 7, 30, 60, 90, 120 – observed cases - FAS

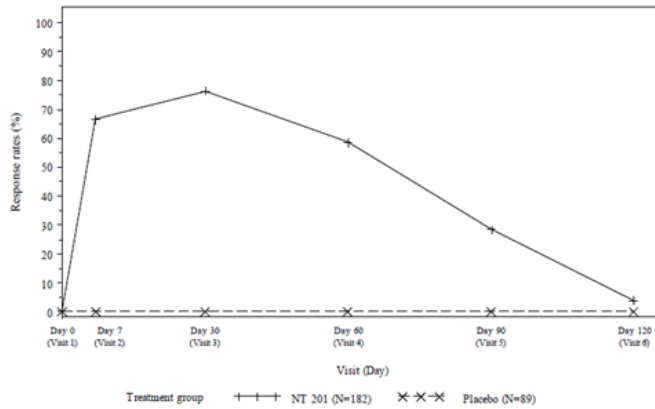


Figure 15: Course over times of percentages of responders at rest by investigator's rating on FWS at Days 7, 30, 60, 90, 120 – observed cases - FAS

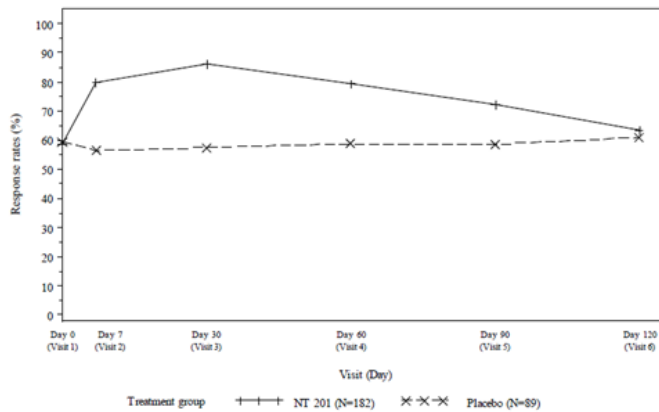
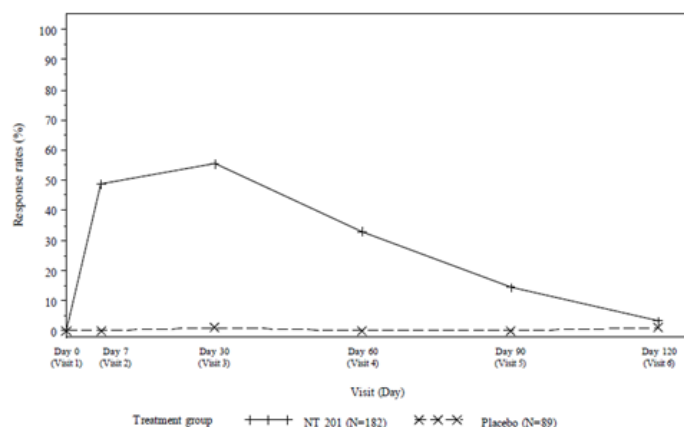


Figure 16: Course over time of percentages of 2-point responders at maximum from by patient's assessment on 4-point scale at Days 7, 30, 60, 90, 120 – observed cases – FAS

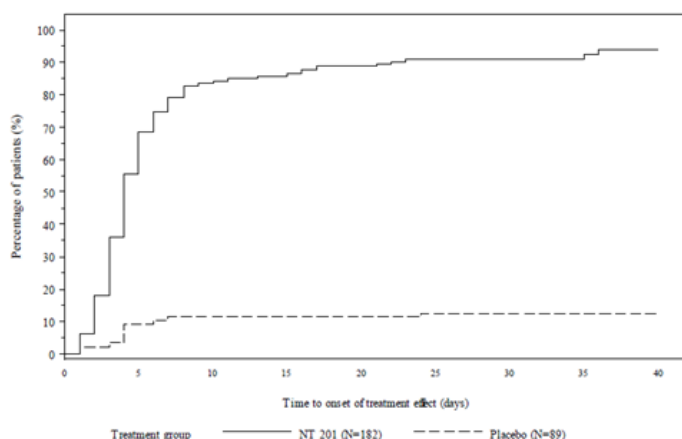


To assess the onset of response with greater temporal resolution, subjects were asked to note the day in which they experienced a subjective onset of a treatment effect. The results of this analysis are shown in the table below. The median day of onset was Day 4 (3 days post-treatment), compared to Day 3 in Study 0724, but the overall temporal profile was similar in the two studies. A small number of subjects reported an onset after 10 days. The cumulative proportion of subjects reporting a response is shown in the subsequent figure.

Table 26: Time to onset of treatment effect – observed cases - FAS

	NT 201 (N=182)	Placebo (N=89)
Onset of treatment effect (recognized by subject until Visit 3)		
Yes n (%)	168 (92.3)	11 (12.4)
No n (%)	14 (7.7)	76 (85.4)
Missing n (%)	0 (0.0)	2 (2.2)
Time to onset of treatment effect (days)		
n	168	11
Nmiss	14	78
Mean	5.4	5.6
SD	5.24	6.34
Min	1	1
Q1	3.0	3.0
Median	4.0	4.0
Q3	6.0	6.0
Max	36	24
Time to onset of treatment effect		
1 day n (%)	11 (6.0)	2 (2.2)
2 days n (%)	22 (12.1)	0 (0.0)
3 days n (%)	33 (18.1)	1 (1.1)
4 days n (%)	35 (19.2)	5 (5.6)
5 days n (%)	24 (13.2)	0 (0.0)
6 days n (%)	11 (6.0)	1 (1.1)
7 days n (%)	8 (4.4)	1 (1.1)
8 days n (%)	6 (3.3)	0 (0.0)
9 days n (%)	2 (1.1)	0 (0.0)
10 days n (%)	1 (0.5)	0 (0.0)
>10 days n (%)	15 (8.2)	1 (1.1)
Missing n (%)	14 (7.7)	78 (87.6)

FAS = Full Analysis Set; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N; SD = Standard deviation.

Figure 17: Time to onset of treatment effect – observed cases - FAS

In support of the validity of the investigator-based FWS assessments, an independent expert committee assessing photos from Day 30 produced similar results. One difference in the two assessments was that a small number of placebo subjects were rated as having a two-point improvement by the expert committee, compared to zero responses in the primary investigator-based analysis. Committee-based results for other definitions of response were similar to those obtained with the investigator-based results (not shown).

Table 27: Percentage of 2-point responders at maximum frown at Day 30 by an independent expert committee according to FWS based on standardised digital photographs – observed cases - FAS

	NT 201 (N=182)	Placebo (N=89)
2-point response		
Yes [n (%)]	117 (64.3)	4 (4.5)
No [n (%)]	64 (35.2)	82 (92.1)
Missing [n (%)]	1 (0.5)	3 (3.4)
Response rate (RR)	0.65	0.05
Difference RR (NT 201-Placebo)	0.60	
95% CI for difference RR	0.51, 0.69	
Odds ratio (OR)	37.48	
Exact 95% CI for OR	13.02, 145.35	
p-value (Fisher's exact test; descriptive)	<0.0001	

CI = Confidence interval; FAS = Full Analysis Set; FWS = Facial Wrinkle Scale; N = Number of subjects in specific group; n = Number of subjects; Calculation of percentages based on N; NA = Not applicable.

Note: Visit 3 = Day 30.

Response rates are calculated based on the number of subjects with available response information. A 2-point responder at maximum frown is a subject with an improvement of at least 2 points on the FWS compared to baseline. The final variable containing the response Yes/No for each subject, visit and situation was generated by applying majority vote on the three derived auxiliary response variables from the experts. That means, the subject (with respective visit and situation) was considered as 2-point responder in case the subject had an improvement of at least 2 points in three or two expert ratings.

7.2.2.14. Subgroup analyses

Subgroup analyses were performed for gender, age, ethnicity, pre-treatment status and baseline disease severity. For analyses with adequate power, the results were statistically significant. For some of the smaller subgroups, such as males (n = 12, CETS response = 25%) and those with only moderate disease at baseline (n = 18, CETS response = 11.1%), the results were not significant, despite a zero response rate in the placebo group. In the case of the moderate subgroup, these results are concordant with Study 0724 and raise the prospect that treatment is less beneficial in the setting of milder disease. The pooled analysis of both studies was also subjected to a subgroup analysis.

7.3. Cervical dystonia

The sponsor submitted two pivotal studies in cervical dystonia (CD), which are discussed below. Study 0408 was a placebo-controlled study, and Study 0013 was an active-controlled, non-inferiority study. Two supportive studies in CD were also submitted.

7.3.1. Pivotal Study 0408

7.3.1.1. Study design, objectives, locations and dates

This Phase III study was titled: 'Prospective, double-blind, placebo-controlled, randomized, multicentre trial with a double-blind parallel-group extension period to investigate the efficacy and safety of different doses of NT 201 in the treatment of cervical dystonia.' It was conducted in 37 centres in the USA, and was conducted from July 2006 to March 2008. The main study period was 21 weeks for each patient and involved a single dose; many patients also entered an extension phase of up to 68 weeks, in which repeated doses were administered as needed. The extension study lasted from October 2006 to June 2009.

The objective of the study was to assess the efficacy and safety of NT 201 at two different doses (240 U and 120 U) in comparison to placebo in the treatment of cervical dystonia in a mixed population of previously treated subjects and at least 40% treatment-naïve subjects.

7.3.1.2. Inclusion and exclusion criteria

Subjects were eligible if they were adults of either gender, aged 18 to 75 years, and had CD of a predominantly rotational form, with sufficient dystonia as rated by the Toronto Western Spasmodic Torticollis Rating Scale (TWSTRS). This scale is described in detail below. It consists of three sub-scores for Severity, Disability and Pain, which are also summed to produce a Total score.

Specifically, subjects needed to satisfy the following criteria:

- TWSTRS-Total score > 20.
- TWSTRS-Severity score > 10.
- TWSTRS-Disability score > 3.
- TWSTRS-Pain score > 1.

Previous treatment with botulinum toxin was permitted, but had to have taken place at least 10 weeks earlier, and the therapeutic response had to have stabilised. The maximum dose of prior treatment was 300 U of type A toxin or 12000 U of type B toxin. Other treatments for dystonia had to be stable for at least three months prior to study entry.

Table 28: Inclusion criteria

Inclusion Criteria	Screening Visit 1	Baseline Visit 2
Male or female outpatients aged 18 to 75 years (inclusive)	X	
A clinical diagnosis of cervical dystonia, i.e. spasmodic torticollis, with predominantly rotational form.	X	
A need for injection (determined by the TWSTRS-Total score)		X
TWSTRS-Total score \geq 20	X	X
TWSTRS-Severity score \geq 10	X	X
TWSTRS-Disability score \geq 3	X	X
TWSTRS-Pain score \geq 1	X	X
On a stable dose of other medications (if any) used for focal dystonia treatment (e.g. anticholinergics and benzodiazepines) for at least 3 months prior to and expected throughout the Main Period	X	X
Negative pregnancy test at trial entry for females of childbearing potential (premenopausal females or postmenopausal females with less than 12 months after last menses)	X (blood)	X (urine)
Women of childbearing potential (premenopausal females or postmenopausal females with less than 12 months after last menses) with reliable means of contraception: hormonal contraception or intrauterine contraceptive device combined with barrier contraception	X	
Willingness of the subject to participate as documented by signed informed consent	X	
<u>For pre-treated subjects only:</u>		
Source documentation of the last two consecutive injection sessions with Botulinum toxin and stable therapeutic response directly prior to trial entry	X	
<u>For pre-treated subjects only:</u>		
At least 10 weeks must have been passed between the last injection with Botulinum toxin for cervical dystonia and Baseline		X
<u>For pre-treated subjects only:</u>		
The most recent injection with Botulinum toxin must have been maximal 300 Units of type A or 12000 Units of type B	X	

Exclusion criteria were as listed below. These were reasonable criteria, aimed at selecting a homogeneous group of well-defined CD subjects with no major confounding features or contraindications to treatment.

Table 29: Inclusion criteria

Exclusion Criteria	Screening Visit 1	Baseline Visit 2
Traumatic torticollis or tardive torticollis	X	
TWSTRS-Severity Score for anterocollis $>$ 2 (pure anterocollis)	X	X
TWSTRS-Severity Score for retrocollis $>$ 2 (pure retrocollis)	X	X
Myotomy or denervation surgery in the affected muscles (e.g. peripheral denervation and/or spinal cord stimulation)	X	
Hypersensitivity to human serum albumin, sucrose or Botulinum toxin Type A	X	
Diagnosis of myasthenia gravis, Lambert-Eaton syndrome, amyotrophic lateral sclerosis or any other significant neuromuscular disease which might interfere with the trial	X	
Current swallowing disorder of any origin (dysphagia scale \geq 3 = severe, with swallowing difficulties and requiring a change in diet)	X	X
Marked limitation on passive range of motion that suggests contractures or other structural abnormality, e.g. cervical contractures or cervical spine syndrome	X	
Treatment with Botulinum toxins for any indication other than cervical dystonia within 4 months prior to Baseline and during the trial	X	X
Known current alcoholism or other drug abuse / dependence	X	
Severe or uncontrolled systemic disease (e.g. cardiac, renal, pulmonary, hepatic, or gastrointestinal), malignant tumor, or known HIV infection in medical history	X	X
Administration of study drug represents a substantial safety concern		X
Present or pre-existing blood coagulation disorders including therapeutic or prophylactic anticoagulation (e.g. heparin, phenprocoumon)	X	X
Infection in the area of the planned injection points or systemic infection presenting a hazard for local injections into the neck muscles	X	X
Non-authorized concomitant treatments	X	X
Planned elective surgery under general anesthesia during the trial	X	
Nursing mothers	X	
Severe impairment of cognition and/or communication that in the opinion of the Investigator does not enable the subject to understand the Informed Consent and participate in the trial	X	
Patients who are currently involved or were involved within the past three months in another therapeutic research study	X	
Previous participation in this clinical study	X	

7.3.1.3. Study treatments

Subjects were randomised to one of the following treatments:

- 240 U of NT 201 in 4.8 mL sterile sodium chloride (NaCl) 0.9% solution;
- 120 U of NT 201 in 4.8 mL sterile NaCl 0.9% solution;
- placebo solution of the same volume (4.8 mL).

The 240 U NT 201 dose corresponds with the median total Botox dose used in the Phase III pivotal study of Botox conducted in the U.S. (Allergan Pharmaceuticals Ltd, 2004). This dose also conforms to standard dosing recommendations in the treatment of cervical dystonia, but in practice the dose is tailored to individual patients. In this study, the number and sites of muscles to be injected were determined by the investigator, according to standard clinical practice for treatment of CD, but the total dose was not individualised. For subjects in the 240 U group, individual muscles were injected with doses comparable to those used for Botox, administered directly into the muscles thought to be responsible for the dystonia. Investigators were blinded to the subject's dose group, so those receiving 120 U received volumes in each muscle that were similar to the 240 U group, but the actual units administered to each muscle were half of what would be considered a standard dose. Placebo subjects were also dosed by volume into individual muscles as though they were receiving a total of 240 U.

7.3.1.4. Efficacy variables and outcomes

The main efficacy variable was a complex scoring system specifically designed to assess the nature and severity of cervical dystonia: the Toronto Western Spasmodic Torticollis Rating Scale (TWSTRS). This scale has three sub-components, Severity, Disability, and Pain, which are in turn derived by combining smaller subcomponents, as described below. This scale seems broadly appropriate for assessment of CD, and it has been discussed and validated in the literature, with acceptable inter-rater reliability and positive correlations between subscores.¹² All raters involved in the study received pre-trial video training in the use of the scale.

Efficacy variables were defined by the sponsor as follows:

Primary efficacy parameter:

Change from baseline in the TWSTRS-Total score at Visit 3 (Week 4)

Secondary efficacy parameters:

Change from baseline in the TWSTRS-Total score at all other post-baseline visits

Change from baseline in TWSTRS-Disability, -Severity and -Pain subscores at all post-baseline visits

Patient Evaluation of Global Response (PEGR) at Visit 5 (Final Visit)

Tertiary efficacy parameters:

Duration of effect (interval from first injection to subsequent injection during the extension phase)

Time to onset and time to waning of treatment effect after injection (subjective estimation)

Global assessment of efficacy by Investigator (4-point Likert scale)

The TWSTRS-Severity Scale was used for verification of eligibility and classification of each subject's head deviation. It consists of the following subdomains:

- Maximal Excursion (Rotation, latero-antero-collis/retro-collis).
- Duration factor.

¹² Bachmann M, Goertelmeyer R, Pesch U, et al. Factorial Validity of the Toronto Western Spasmodic Torticollis Rating Scale (TWSTRS). *Mov Disord* 2000;15(Suppl 3):140; Comella CL, Leurgans S, Stebbins GT. Factor Analysis and Internal Consistency of the Toronto Western Spasmodic Torticollis Rating Scale (TWSTRS). *Mov Disord* 1999;14(5):903; Consky E, Basinski A, Belle L, et al. The Toronto Western Spasmodic Torticollis Rating Scale (TWSTRS): Assessment of Validity and Inter-Rater Reliability. *Neurology* 1990;40(Suppl 1):1199P.

- Effect of sensory tricks.
- Shoulder elevation.
- Range of motion.
- Time factor.

The TWSTRS-Severity scale ranges from 0 to 35, with higher scores indicating greater impairment.

The TWSTRS-Disability Scale attempts to assess the degree to which daily activities are affected by cervical dystonia. The subject is asked questions relating to work performance, feeding, dressing, hygiene, driving, reading, television viewing, and leisure activities outside the home. Each category is reported on a 6-point scale and the total disability score ranges from 0 to 30 points, with higher scores indicating greater impairment.

The TWSTRS-Pain Scale was used for the verification of eligibility and as a minor endpoint. The TWSTRS-Pain scale ranges from 0 to 20 points. It is the sum of a pain severity component, a duration component and an assessment of the contribution of pain to disability. The pain severity component ranges from 0 to 10 points, with '0' meaning 'no pain' and '10' the 'most excruciating pain imaginable'. The subject also rates the severity of neck pain experienced during the previous week, indicating the severity when pain was at its 'best', 'worst' and 'usual' status. The pain severity component is then calculated as: $Severity = ((2 \times usual) + best + worst)/4$. For the duration component of the TWSTRS-Pain scale, the subject rates the duration of neck pain on a 6-point scale, and for assessment of the contribution of pain to disability, the subject rates the contribution on a 6-point scale.

Subjects also completed a subjective assessment of their global response to treatment on a 9-point scale, the Patient Evaluation of Global Response (PEGR). This scale ranges from -4 to +4: very marked worsening (- 4), marked worsening (- 3), moderate worsening (- 2), slight worsening (- 1), unchanged (0), slight improvement (+1), moderate improvement (+ 2), marked improvement (+ 3), complete abolishment of signs and symptoms (+ 4). The scale was adapted from Wissel et al (2000),¹³ and seems broadly appropriate for a subjective response scale. The sponsor did not specifically discuss validation of this scale, but it seems fairly standard for a subjective global response scale. It was considered a secondary efficacy variable.

7.3.1.4.1. Global Assessment of Efficacy.

The investigator also performed a global assessment of the efficacy of study treatment at the subject's final visit, using a standard 4-point Likert scale: 'very good', 'good', 'moderate' and 'poor'. This was considered a tertiary efficacy variable.

7.3.1.5. Randomisation and blinding methods

Randomisation was performed with equal likelihood to each treatment group, using a computerised randomisation system (RANCODE, version 3.6). Blinding was maintained by using vials of identical appearance in all treatment groups and by keeping the randomisation code sealed and secured until all data was collected.

7.3.1.6. Analysis populations

The sponsor defined four analysis populations:

- Evaluable for Safety (EFS) Population, consisting of all subjects who received any amount of trial medication during the main period of the study.

¹³ Wissel J, Müller J, Dressnandt J, et al. 2000. Management of spasticity associated pain with botulinum toxin A. J Pain Symptom Manage 20(1):44-49.

- Intent-To-Treat (ITT) Population, consisting of all randomised subjects.
- Treated-Per-Protocol (TPP) Population, consisting of randomised subjects for whom no major protocol deviations were identified.
- Treatment-naïve population, consisting of randomised subjects who had not received any prior treatment with botulinum toxin of any serotype.

The primary efficacy analysis was performed on the ITT population.

7.3.1.7. Statistical methods

The primary efficacy endpoint was the change from Baseline (Visit 2, Day 0) to Visit 3 (Week 4) in the TWSTRS-Total score and the primary analysis was based on the difference between the treatment groups in the ITT population for this endpoint. In particular, the confirmatory analysis was based on the comparison of least square (LS) means from an analysis of covariance (ANCOVA) model at Visit 3 (Week 4) between treatment groups.

The dependent variable in the ANCOVA was the change from baseline of the TWSTRS-Total score and the independent variables were: treatment, baseline TWSTRS-Total score, gender, age, pre-treatment of cervical dystonia, and centre. Because multiple endpoints were assessed in addition to the single primary endpoint, comparisons between treatment groups were performed by using a fixed-sequence test procedure (step downward) in the ITT Population, and no type-I error adjustment was considered necessary. All comparative tests were performed two-sided with a type-I error of $\alpha = 0.05$. Missing data were conservatively replaced with a change score of '0' (no change), which would have the effect of reducing any apparent treatment effect. Sensitivity analyses were also performed, in which the analysis was restricted to observed cases, or to subjects in the PP population.

A similar ANCOVA was performed with results from Visit 4 and the final visit, but this was considered a secondary analysis. TWSTRS subscores were also analysed with ANCOVA, using similar methodology. The secondary endpoint of PEGR was analysed with ANCOVA, and missing data were imputed as zero (no effect). Other endpoints were presented with descriptive statistics. Overall, these statistical methods were appropriate, and the sponsor's interpretation of the results did not appear to depend critically on the specific choice of methods.

7.3.1.8. Sample size

Sample-size estimations were based on a predicted mean change from baseline in TWSTRS-Total score of four points for placebo and of 10 points for 240 U group, with a common standard deviation of nine points, based on previous studies (Brashear et al. 1999a,¹⁴ Using a 2-sided t-test at 5% significance level, it was estimated that a study with N = 49 completing subjects per group would have 90% power to detect a treatment difference between 240 U group and placebo. The sample size was increased to 59 subjects per group to increase the power for comparing the 120 U group and placebo, and then increased again to allow for a 20% dropout rate, to give a final target of 74 subjects per treatment group, with a total sample size of 222 subjects.

¹⁴ Brashear A, Zafonte R, Corcoran M, Galvez-Jimenez N, Gracies JM, Gordon MF, et al. 2002. Inter- and intrarater reliability of the Ashworth Scale and the Disability Assessment Scale in patients with upper-limb poststroke spasticity. *Arch Phys Med Rehabil* 83(10): 1349-1354. (ID 1905248); Brin MF, Lew MF, Adler CH, et al. 1999. Safety and efficacy of NeuroBloc (botulinum toxin type B) in type A-resistant cervical dystonia. *Neurology* 1999;53(7):1431-1438; Truong D, Lew M, Orlova O. 2008. Efficacy of Dysport (Botulinum toxin type A) for the treatment of cervical dystonia: A multicenter, randomized, double-blind, placebo-controlled phase III study. *Toxicon* 51(1):37-38. Toxins 2008 Baveno, Italy.

7.3.1.9. Participant flow

Subject disposition is summarised in the figure and table below. Dropouts in each group were relatively infrequent (three to six subjects per group), and are unlikely to have produced significant bias. The reasons for the withdrawals differed across groups, with lack of efficacy being the most common cause of withdrawal in the placebo group.

Figure 18: Subject disposition (all enrolled subjects) in the main period

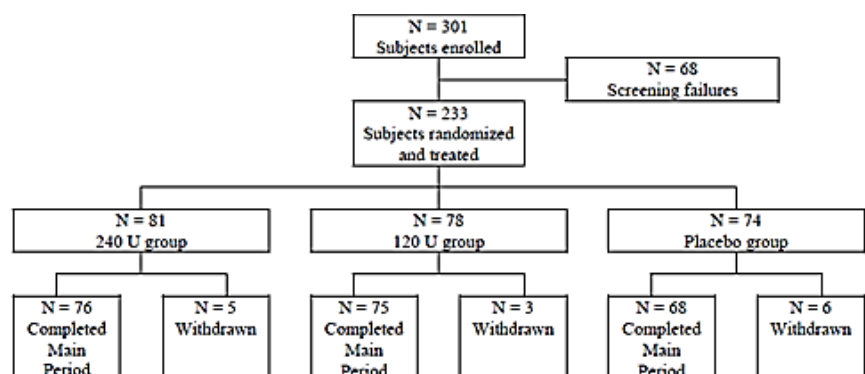


Table 30: Subjects who discontinued the study prematurely during the MP (all randomised subjects, N = 233)

	240 U group	120 U group	Placebo group	Total
Total	N = 81	N = 78	N = 74	N = 233
No. (%) of subjects who completed the main period	76 (93.8)	75 (96.2)	68 (91.9)	219 (94.0)
No. (%) of subjects who did not complete the main period	5 (6.2)	3 (3.8)	6 (8.1)	14 (6.0)
Main reason for premature termination				
Withdrawal criteria occurred	0 (0.0)	0 (0.0)	1 (1.4)	1 (0.4)
Insufficient efficacy	0 (0.0)	0 (0.0)	3 (4.1)	3 (1.3)
Adverse event	2 (2.5)	1 (1.3)	0 (0.0)	3 (1.3)
Consent withdrawn	1 (1.2)	1 (1.3)	1 (1.4)	3 (1.3)
Lost to follow-up	1 (1.2)	1 (1.3)	1 (1.4)	3 (1.3)
Other	1 (1.2)	0 (0.0)	0 (0.0)	1 (0.4)

7.3.1.10. Major protocol violations/deviations

Most subjects (92.3%) entered the TPP population. Protocol violations were relatively rare, and were acceptable for a study of this nature. The violations are summarised below.

Table 31: Major protocol deviations (all randomised subjects, N = 233)

Major Protocol Deviation [a]	240 U group	120 U group	Placebo group	Total
	N = 81	N = 78	N = 74	N = 233
	n (%)	n (%)	n (%)	n (%)
Total subjects with at least one deviation	4 (4.9)	8 (10.3)	6 (8.1)	18 (7.7)
Administration of study medication	1 (1.2)	4 (5.1)	2 (2.7)	7 (3.0)
Different rater	1 (1.2)	3 (3.8)	1 (1.4)	5 (2.1)
Trained rater	1 (1.2)	1 (1.3)	1 (1.4)	3 (1.3)
Early termination	1 (1.2)	0	1 (1.4)	2 (0.9)
Exclusion criterion #14 (#9) not met at Screening (Baseline)	1 (1.2)	0	1 (1.4)	2 (0.9)
Inclusion criterion #1 not met at Baseline	0	1 (1.3)	0	1 (0.4)
Inclusion criterion #4 (#3) not met at Screening (Baseline)	0	1 (1.3)	0	1 (0.4)
Too early re-injection	0	0	1 (1.4)	1 (0.4)

Exclusion criterion #14 (#9) at Screening (Baseline): Non-authorized concomitant treatments
 Inclusion criterion #1 at Baseline: A need for injection (determined by the TWSTRS-Total score)
 Inclusion criterion #4 (#3) at Screening (Baseline): TWSTRS-Severity score ≥ 10
 [a] Not mutually exclusive, i.e. subjects may have had more than one deviation
 U=Units

7.3.1.11. Baseline data

Baseline demographic data and disease characteristics are summarised in the tables below. No important differences existed between the groups at baseline.

Table 32: Demographic data (Total ITT population)

Parameter	240 U group N = 81 n (%)	120 U group N = 78 n (%)	Placebo group N = 74 n (%)	Total N = 233 n (%)
Gender				
Male	27 (33.3%)	27 (34.6%)	25 (33.8%)	79 (33.9%)
Female	54 (66.7%)	51 (65.4%)	49 (66.2%)	154 (66.1%)
Race				
American Indian or Alaskan Native	0 (0.0%)	1 (1.3%)	0 (0.0%)	1 (0.4%)
Asian	0 (0.0%)	0 (0.0%)	3 (4.1%)	3 (1.3%)
Black or African American	2 (2.5%)	0 (0.0%)	4 (5.4%)	6 (2.6%)
Hispanic or Latino	5 (6.2%)	3 (3.8%)	2 (2.7%)	10 (4.3%)
White	74 (91.4%)	74 (94.9%)	65 (87.8%)	213 (91.4%)
Smoking				
Non-smoker	39 (48.1%)	36 (46.2%)	38 (51.4%)	113 (48.5%)
Smoker	22 (27.2%)	19 (24.4%)	11 (14.9%)	52 (22.3%)
Ex-smoker	20 (24.7%)	23 (29.5%)	25 (33.8%)	68 (29.2%)
Age (years)				
Mean (SD)	53.2 (12.17)	52.8 (11.53)	52.4 (10.75)	52.8 (11.47)
Median	53.0	54.0	52.0	54.0
Height (in)				
Mean (SD)	65.9 (3.44)	66.8 (3.67)	66.56 (3.33)	66.4 (3.49)
Median	66.0	66.4	66.0	66.0
Weight (lb)				
Mean (SD)	160.8 (35.46)	178.6 (36.26)	170.1 (35.52)	169.7 (36.35)
Median	160.0	179.0	165.0	165.0
Body Mass Index [a]				
Mean (SD)	25.9 (5.19)	28.1 (5.63)	26.9 (4.49)	27.0 (5.20)
Median	24.4	27.7	26.9	26.4

Source: Table 14.1.3.1.1

[a] BMI was calculated according to the formula: $BMI = (Weight(lb) / Height(in)^2) \times 703$.

SD=Standard Deviation, U=Units

Table 33: Demographic data for subgroups of treatment naïve and pre-treated subjects (ITT population)

Parameter	240 U group		120 U group		Placebo group	
	Treatment-Naïve N=31	Pre-treated N=50	Treatment-Naïve N=31	Pre-treated N=47	Treatment-Naïve N=28	Pre-treated N=46
Gender						
Male	8 (25.8%)	19 (38.0%)	10 (32.3%)	17 (36.2%)	11 (39.3%)	14 (30.4%)
Female	23 (74.2%)	31 (62.0%)	21 (67.7%)	30 (63.8%)	17 (60.7%)	32 (69.6%)
Race						
American Indian or Alaskan Native	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (2.1%)	0 (0.0%)	0 (0.0%)
Asian	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (6.5%)
Black or African American	1 (3.2%)	1 (2.0%)	0 (0.0%)	0 (0.0%)	2 (7.1%)	2 (4.3%)
Hispanic or Latino	2 (6.5%)	3 (6.0%)	0 (0.0%)	3 (6.4%)	2 (7.1%)	0 (0.0%)
White	28 (90.3%)	46 (92.0%)	31 (100.0%)	43 (91.5%)	24 (85.7%)	41 (89.1%)
Smoking						
Non-smoker	15 (48.4%)	24 (48.0%)	17 (54.8%)	19 (40.4%)	16 (57.1%)	22 (47.8%)
Smoker	9 (29.0%)	13 (26.0%)	6 (19.4%)	13 (27.7%)	6 (21.4%)	5 (10.9%)
Ex-smoker	7 (22.6%)	13 (26.0%)	8 (25.8%)	15 (31.9%)	6 (21.4%)	19 (41.3%)
Age (years)						
Mean (SD)	50.9 (12.07)	54.7 (12.13)	50.9 (11.89)	54.0 (11.23)	51.5 (10.51)	53.0 (10.97)
Median	51.0	55.5	53.0	57.0	53.0	52.0
Height (in)						
Mean (SD)	65.4 (3.13)	66.2 (3.61)	67.2 (3.39)	66.6 (3.86)	66.9 (3.09)	66.4 (3.48)
Median	65.0	66.0	66.8	66.0	66.0	66.5
Weight (lb)						
Mean (SD)	161.2 (39.96)	160.5 (32.67)	164.9 (30.67)	187.6 (37.10)	173.9 (37.97)	167.9 (34.28)
Median	158.1	161.0	161.9	191.0	165.0	164.0
Body Mass Index [a]						
Mean (SD)	26.5 (6.43)	25.6 (4.25)	25.7 (4.58)	29.7 (5.72)	27.1 (4.45)	26.7 (4.56)
Median	24.7	24.3	24.7	28.7	27.4	26.8

Source: Table 14.1.3.1.1

[a] BMI was calculated according to the formula: $BMI = (Weight(lb) / Height(in)^2) \times 703$.

SD=Standard Deviation, U=Units

Table 34: Medical history by MedDRA SOC (applying to > 5% of all subjects) and PT (applying to > 3% of all subjects) (Total ITT population)

System Organ Class Preferred Term [a]	240 U group N=81 n (%)	120 U group N=78 n (%)	Placebo group N=74 n (%)	Total N=233 n (%)
Subjects with medical history	65 (80.2)	57 (73.1)	54 (73.0)	176 (75.5)
Surgical and medical procedures	55 (67.9)	46 (59.0)	41 (55.4)	142 (60.9)
Hysterectomy	13 (16.0)	8 (10.3)	11 (14.9)	32 (13.7)
Tubal ligation	5 (6.2)	12 (15.4)	9 (12.2)	26 (11.2)
Cholecystectomy	4 (4.9)	8 (10.3)	6 (8.1)	18 (7.7)
Tonsillectomy	6 (7.4)	6 (7.7)	5 (6.8)	17 (7.3)
Appendectomy	4 (4.9)	6 (7.7)	7 (9.5)	17 (7.3)
Intervertebral disc operation	2 (2.5)	3 (3.8)	4 (5.4)	9 (3.9)
Infections and infestations	10 (12.3)	12 (15.4)	8 (10.8)	30 (12.9)
Injury, poisoning and procedural complications	15 (18.5)	9 (11.5)	6 (8.1)	30 (12.9)
Neoplasms benign, malignant and unspecified (incl. cysts and polyps)	13 (16.0)	4 (5.1)	9 (12.2)	26 (11.2)
Gastrointestinal disorders	9 (11.1)	5 (6.4)	7 (9.5)	21 (9.0)
Nervous system disorders	7 (8.6)	4 (5.1)	7 (9.5)	18 (7.7)
Psychiatric disorders	5 (6.2)	5 (6.4)	4 (5.4)	14 (6.0)
Investigations	3 (3.7)	5 (6.4)	6 (8.1)	14 (6.0)
Musculoskeletal and connective tissue disorders	4 (4.9)	6 (7.7)	2 (2.7)	12 (5.2)

Source: Table 14.1.3.6.1

[a] Medical History was coded by using MedDRA version 9.1.

U=Units

Table 35: History of cervical dystonia (total ITT population)

	240 U group N=81	120 U group N=78	Placebo group N=74	Total N=233
Observable Displacement of Head Position (n, %)				
Yes	81 (100.0)	78 (100.0)	74 (100.0)	233 (100.0)
No	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Duration Since First Diagnosis (months)				
Mean (SD)	49.4 (67.18)	52.4 (66.49)	53.9 (72.16)	51.9 (68.18)
Median	17.6	24.7	22.4	19.5
Estimated Duration of CD (months)				
Mean (SD)	116.7 (107.38)	111.0 (100.48)	129.7 (107.72)	119.0 (105.08)
Median	79.0	72.0	120.0	84.0
Surgery for CD (n, %)				
Yes	1 (1.2)	0 (0.0)	0 (0.0)	1 (0.4)
No	80 (98.8)	78 (100.0)	74 (100.0)	232 (99.6)
Muscle Hypertrophy (n, %)				
Yes	59 (72.8)	54 (69.2)	47 (63.5)	160 (68.7)
No	22 (27.2)	24 (30.8)	27 (36.5)	73 (31.3)
CD Type (n, %)				
Rotational - Right	41 (50.6)	44 (56.4)	31 (41.9)	116 (49.8)
Rotational - Left	39 (48.1)	33 (42.3)	42 (56.8)	114 (48.9)
Rotational - Total	80 (98.8)	77 (98.7)	73 (98.6)	230 (98.7)
Retrocollis	20 (24.7)	18 (23.1)	20 (27.0)	58 (24.9)
Anterocollis	18 (22.2)	18 (23.1)	13 (17.6)	49 (21.0)
Laterocollis - Right	27 (33.3)	35 (44.9)	23 (31.1)	85 (36.5)
Laterocollis - Left	37 (45.7)	18 (23.1)	22 (29.7)	77 (33.0)
Laterocollis - Total	64 (79.0)	53 (67.9)	45 (60.8)	162 (69.5)
Shift - Lateral	21 (25.9)	19 (24.4)	18 (24.3)	58 (24.9)
Shift - Sagittal	6 (7.4)	13 (16.7)	6 (8.1)	25 (10.7)
Shift - Lateral and Sagittal	0 (0.0)	0 (0.0)	1 (1.4)	1 (0.4)
Shift - Total	27 (33.3)	32 (41.0)	25 (33.8)	84 (36.1)
Shoulder Elevation	57 (70.4)	59 (75.6)	56 (75.7)	172 (73.8)
Other	2 (2.5)	3 (3.8)	1 (1.4)	6 (2.6)
Relative Suffering from CD (n, %)				
Yes	6 (7.4)	11 (14.1)	4 (5.4)	21 (9.0)
No	75 (92.6)	67 (85.9)	70 (94.6)	212 (91.0)

Source: Table 14.1.3.2

CD=Cervical Dystonia, U=Units

Table 36: Duration since first diagnosis and estimated duration of cervical dystonia in subgroups of treatment-naïve and pre-treated subjects (ITT population)

Parameter	240 U group		120 U group		Placebo group	
	Treatment-Naïve N=31	Pre-treated N=50	Treatment-Naïve N=31	Pre-treated N=47	Treatment-Naïve N=28	Pre-treated N=46
Duration Since First Diagnosis (months)						
Mean (SD)	8.7 (18.85)	84.9 (73.97)	25.3 (59.40)	73.5 (64.73)	5.4 (16.69)	93.0 (76.04)
Median	2.1	62.0	1.1	53.8	0.8	62.8
Estimated Duration of CD (months)						
Mean (SD)	88.5 (108.62)	134.1 (103.89)	113.6 (118.03)	109.4 (88.83)	106.8 (116.34)	143.7 (100.88)
Median	60.0	102.0	72.0	92.0	60.0	122.5

Source: Table 14.1.3.2
CD=Cervical Dystonia, SD=Standard Deviation, U=Units

7.3.1.12. Results for the primary efficacy outcome

The raw results are tabulated below for the total ITT population, and for the treatment-naïve and pre-treated subgroups. The third table below shows the between-group differences based on a least squares (LS) approach, with the corresponding 95% CIs and p-values. A minor improvement was seen in the placebo group, but a more substantial improvement was seen with active treatment. The differences between each dose and placebo were statistically significant ($p < 0.001$) by ANCOVA. There was a trend in favour of the higher dose over the lower dose, but the improvement was only 1.0 point better in the higher dose group, from a baseline TWSTRS-Total of approximately 2 points, and this difference was not significant. A range of sensitivity analyses, using observed cases instead of imputation, or using the full ANCOVA model or reduced models with fewer factors, all produced very similar conclusions (not shown).

Table 37: Mean (SD) TWSTRS-total score at baseline in total ITT population, treatment-naïve and pre-treated subjects

(Sub)Group	240 U group	120 U group	Placebo group
Total ITT population	42.1(9.30) (N=81)	42.6(9.67) (N=78)	41.8(7.87) (N=74)
Treatment-Naïve Subjects	40.1(9.19) (N=31)	41.9(9.71) (N=31)	41.3(6.52) (N=28)
Pre-Treated Subjects	43.4(9.25) (N=50)	43.1(9.71) (N=47)	42.0(8.65) (N=46)

Table 38: Mean change from baseline to Visit 3 in TWSTRS-total score in total ITT population. Treatment-naïve and pre-treated subjects (missing values replaced by the subject's baseline value).

(Sub)Group	240 U group	120 U group	Placebo group
Total ITT population	-10.9(11.72) (N=81)	-9.9(10.35) (N=78)	-2.2(7.30) (N=74)
Treatment-Naïve Subjects	-10.0(9.16) (N=31)	-11.9(11.12) (N=31)	-2.0(5.97) (N=28)
Pre-Treated Subjects	-11.44(13.12) (N=50)	-8.53(9.71) (N=47)	-2.4(8.07) (N=46)

Source: Tables 14.2.1.1.6.1
U=Units

Table 39: LS mean treatment difference in change from baseline to Visit 3 in TWSTRS-total score – total ITT population, treatment-naïve and pre-treated subjects

Comparison	Subgroup	LS mean treatment difference [95% CI]	p-value "Treatment" (Full model)
240 U vs. Placebo group	Treatment-Naïve Subjects	-7.5 [-11.6;-3.5]	<0.001
	Pre-Treated Subjects	-10.1 [-14.4;-5.9]	<0.001
120 U vs. Placebo group	Treatment-Naïve Subjects	-9.6 [-14.2;-5.1]	<0.001
	Pre-Treated Subjects	-5.7 [-9.3;-2.1]	0.002
240 U vs. 120 U group	Treatment-Naïve Subjects	-2.1 [-7.1;2.9]	0.405
	Pre-Treated Subjects	3.4 [-1.1;7.9]	0.135

Source: Table 14.2.1.2.1.1

N=number of patients, LS=Least Square, CI=Confidence Interval; U=Units

7.3.1.13. Results for other efficacy outcomes

The primary analysis was based on results from Visit 3 (Week 4), and results from other time points were considered secondary. The Week 8 results resembled the Week 4 results, but there was a minor waning of efficacy. By the final visit, the change from baseline had lessened again.

Results for the individual subscores of TWSTRS (Severity, Disability and Pain) were qualitatively similar to those obtained for the total TWSTRS. All comparisons between active treatment and placebo were statistically significant. For severity scores at Visit 3, the difference between the high and low dose was also marginally significant, in favour of the higher dose.

The PEGR showed that most subjects receiving active treatment felt they had improved, whereas most placebo subjects felt they were unchanged or worse. Results in the 240 U group were slightly better than the 120 U group.

The time to onset of subjective treatment effect was 6-7 days across all groups, even in the placebo group. Waning was noted after 7-8 weeks in the active groups. The duration of treatment effect, based on the time to re-treatment, was 10-12 weeks, similar to other studies.

By the Mann-Whitney test, investigator assessments of therapeutic efficacy were significantly better in both active groups in comparison to placebo ($p < 0.001$ for both dose groups).

7.3.1.14. Subgroup analyses

Subgroup analyses revealed that gender and baseline TWSTRS score had a significant effect on the full ANCOVA model comparison of NT 201 and placebo. Males showed a weaker response to treatment, possibly reflecting a higher muscle mass. Subjects with high baseline scores showed larger reductions in score with treatment, which may reflect that they had more points to lose and more dystonia in need of treatment. Age and pre-treatment status did not have a significant effect on the ANCOVA model.

7.3.2. Pivotal Study 0013

7.3.2.1. Study design, objectives, locations and dates

This Phase III study was titled: "Safety and efficacy of NT 201 (highly purified Botulinum Neurotoxin A) compared to Botox (purified Botulinum Neurotoxin A-complex) in cervical dystonia." It was conducted in 51 centres located in Belgium, the Czech Republic, France, Russia, Germany, Slovakia, Sweden, Austria, Poland, Hungary and Israel, from March 2001 to April 2002.

It was a non-inferiority study, using a double-blind, active control, parallel-group design. The objective was to show that a single dose of NT 201 (individualised for each patient) was not inferior to a matching dose of Botox in the treatment of CD, in terms of efficacy and safety. The study was performed in subjects who had previously been treated with Botox for the same condition.

7.3.2.2. Inclusion and exclusion criteria

Eligible subjects were adults of either gender aged less than or equal to 75 years, with predominantly rotational cervical dystonia, who had not received Botox for at least 10 weeks. They had to have a TWSTRS–Severity score greater than or equal to 10, a score for rotation greater than or equal to 2, and a score for rotation that was higher than the score for laterocollis, and antero/retrocollis. They had to have had a stable therapeutic response to Botox with their previous two treatments, and the Botox dose had to be in the range 70-300 U. Other medications used to treat dystonia had to have been kept at stable doses for greater than or equal to 3 months, and, if female, they needed to be able to avoid pregnancy.

Exclusion criteria were standard, and similar to the other pivotal CD study (Study 0408, described above). Patients were not eligible if they had received botulinum toxin preparations other than Botox in the last two treatments, or if they had undergone myotomy or previous denervation. This is reasonable as such prior treatments could confound the results.

7.3.2.3. Study treatments

All subjects received a dose of 70-300 U of NT 201 or Botox, diluted in normal saline, and administered intramuscularly to the neck muscles deemed responsible for the dystonia. The number and sites of injection was individualised, based on the patients' previous treatment with Botox.

Subjects only received a single dose, and were then followed for up to 16 weeks.

7.3.3. Efficacy variables and outcomes

The main efficacy variable was the TWSTRS score, as described for Study 0408. The primary endpoint was the change in the TWSTRS–Severity subscore from baseline to the “control visit” at Day 28 plus or minus 3. Subjects were also assessed at a final visit on Day 109-112.

Secondary endpoints included:

- TWSTRS-Pain subscore;
- Visual Analogue Score (VAS) for pain;
- Treatment duration, defined as the time from injection to the visit in which the TWSTRS–Severity had returned to at least 80% of baseline;
- Response rate, defined as the proportion of subjects with an improvement greater than or equal to 20% in their TWSTRS–Severity score at the control visit;
- TWSTRS-Factorial scores (see below);
- Time to onset and time to waning of treatment effect, as judged subjectively by each patient;
- Investigators Global Assessment of efficacy, based on a 4-point rating scale (“very good”, “good”, “moderate”, “poor”).

The standard subscores of the TWSTRS score were described for Study 0408. In addition to the standard subscores, this study assessed “TWSTRS-Factorial Scores”, by following the method of Bachmann et al (200). This method groups subcomponents of the TWSTRS–Severity score into 3 domains or “factors”:

- Factor 1, Motion and Duration.
- Rotation, Duration, Motion.
- Time spent maintaining the neutral position.
- Factor 2, Position.
- Laterocollis, Antero/Retrocollis.

- Shoulder Elevation.
- Factor3, Effect of sensory tricks.

This factor analysis was considered exploratory, and merely sought to clarify which aspects of the score were responsive to treatment.

7.3.3.1. Randomisation and blinding methods

Subjects were randomised in equal proportions to Botox and NT 201, using RANCODE via a centralised randomisation procedure.

Blinding was maintained by using identical syringes for both study treatments, prepared offsite, and by keeping randomisation codes sealed and secure. Unblinding was very unlikely given that the two agents share the same mechanism of action and Phase I studies suggested similar potency per unit.

7.3.3.2. Analysis populations

The sponsor defined three study populations:

- Evaluation for Safety Sample (EFS), consisting of all randomised subjects.
- Intent-to-Treat Sample (ITT), consisting of all patients who were randomised and received trial medication.
- Treated Per-Protocol Sample (TPP), consisting of the ITT subjects without major protocol violations.

The primary analysis was based on the TPP population, which is noteworthy. Pivotal Phase III studies are usually based on the ITT population, and TGA guidelines encourage the use of intent-to-treat analyses. One of the main reasons for preferring ITT analyses is that protocol violations may reflect poor compliance or patient withdrawals that arise from side effects or other practical issues likely to be encountered in real clinical practice. These protocol violations might be non-randomly distributed in the treatment groups because, for instance, active treatment causes side effects and hence dropouts, and therefore it is usually more appropriate to consider the intention to treat, before there has been a chance for such issues to arise.

On the other hand, some protocol violations (such as misdosing or recruitment of ineligible subjects) may add noise to the data, reducing the power of a study - statistical significance is often more marked in a per-protocol analysis. This means that a negative (non-significant) result in an ITT analysis might be partially due to protocol violations. For a study intended to produce a negative result, such as a non-inferiority study, it may therefore be more conservative to consider the per-protocol analysis. Also, concerns about potential withdrawal bias are less applicable to a single-dose injected treatment, because subjects were not able to withdraw from the continued effects of the injection once they had received it.

The analysis of the TPP population is therefore broadly acceptable, but it would be of concern if the results of the ITT and TPP analyses were discordant.

7.3.3.3. Statistical methods

The hypothesis being tested was that NT 201 was not worse than Botox, as reflected in the TWSTRS-Severity score at Day 28 plus or minus 3. The primary analysis method was an ANCOVA, based on the TPP population. The dependent variable was change in TWSTRS-Severity score. The independent variables were treatment, baseline TWSTRS-Severity, total dose, gender, age, number of injections since diagnosis of torticollis, country, and country*treatment interaction. The difference between treatments was to be expressed with a least-square (LS) means approach.

A score difference of less than 1.3 points was considered clinically irrelevant. The reason for choosing this particular difference was that it represented half of the observed placebo-

difference (2.6 points) previously cited for TWSTRS-Severity scores in the published literature.¹⁵ A one-sided alpha value of 0.025 was used to assess for a significant difference, with a power of 90% (see the discussion of sample size below).

The sponsor specified that three ANCOVA models were to be presented, and stated that the “final model” was the one to be applied on confirmatory testing:

- The full model including all variables and covariates;
- the final model including only those adjusting variables that in a backward selection procedure presented a p less than or equal to 0.2 ; and
- the simple model including the treatment effect only.

Secondary endpoints were also based on the TPP analysis, and were expressed with exploratory and descriptive statistic only.

7.3.3.4. Sample size

Sample size was initially estimated from a blinded interim analysis of 107 patients included in the trial. An ANCOVA at that point revealed a common standard deviation for change in TWSTRS-Severity score of 3.51, and suggested that 310 TPP patients would be needed to demonstrate that the difference between groups was smaller than 1.3 points, with a one-sided alpha level of 0.025 and power of 90%. Assuming approximately 10% protocol violations, it was estimated that 346 patients would need to be randomised to show non-inferiority of NT 201 versus Botox.

These estimates appeared satisfactory, but further interim analyses suggested that a larger study would be needed, and the protocol was amended to have a target recruitment of 460 patients. The final study was adequately powered, as evident in the narrow estimates for the treatment difference that emerged from the final ANCOVA (see the results section, below).

7.3.3.5. Participant flow

Of note, the ITT population included 463 patients, whereas the TPP population only had 420 patients, because of 43 major protocol deviations. Of the 463 randomised subjects, most (451) completed the study; reasons for discontinuation are shown in the table below. Note that, in most cases, protocol deviations did not lead to withdrawal from the study. The nature of the protocol deviations is discussed in the following section.

¹⁵ Brashear A, Gordon MF, Elovic E, et al. 2002. Intramuscular injection of botulinum toxin for the treatment of wrist and finger spasticity after a stroke. *N Engl J Med* 347(6):395-400; Brin MF, Lew MF, Adler CH, et al. 1999. Safety and efficacy of NeuroBloc (botulinum toxin type B) in type A-resistant cervical dystonia. *Neurology* 53(7):1431-1438.

Table 40. Patients who discontinued therapy or study prematurely (ITT; N = 463)

	Treatments; Number of Patients (%) ^a		
	NT 291 (N = 231)	BTXCo (N = 232)	Total (N = 463)
Completed Study	226 (97.8)	225 (97.0)	451 (97.4)
Discontinued Study	5 (2.2)	7 (3.0)	12 (2.6)
Withdrawal Reason			
Lack of efficacy (under "main reasons")	3 (1.3); Nos. S117/125, S139/163, S499/1124	2 (0.9); Nos. S138/162, S501/1126	5 (1.1)
Control visit not performed (under "other reasons")	-	4 (1.7); Nos. S094/052, S098/068, S243/272, S540/1185	4 (0.9)
Lost to follow-up (under "main reasons")	1 (0.4); No. S261/285	1 (0.4); No. S549/483	2 (0.4)
Protocol deviation (under "other reasons")	2 (0.9); Nos. S009/065, S117/125	-	2 (0.4)
Adverse event (under "main reasons")	-	1 (0.4); S540/1185	1 (0.2)

^a Percentages based on the total number of patients in a patient group.
 Data Source: Tables 14.1.2.1.9.1 – 9 and 16.2.1.1.9.
 Abbreviations: N = number of patients.
 Note that multiple reasons for discontinuation may have been reported for a single patient.

7.3.3.6. Major protocol violations/deviations

Primary analysis of this study was based on the TPP population, rather than on the full ITT population. Reasons for excluding subjects from the TPP population are listed below. The overall number of violations is reasonable for a study of this size and complexity, but it is of concern that so many patients were censored from the sponsor's final analysis.

In many cases, excluded subjects did not satisfy the inclusion criteria, so they should not have been in the study in the first place, and their exclusion from the primary analysis (based on the TPP population) is reasonable. For some cases, control visits were not conducted at all, or were not conducted at the appropriate time, so their TWSTRS scores were either missing or were potentially not reflective of the efficacy of the study treatment at the intended time of assessment. Thus, their exclusion from the TPP analysis is likely to have improved accuracy of the comparison between the two treatments.

Protocol violations were, at least, balanced across the treatment groups, so the TPP population is not likely to have been substantially affected by withdrawal bias.

Table 41. Major Protocol Deviations (ITT, N = 463)

Type of violation	Treatments; Number of patients (%)(No.) ^a		
	NT 201 (N = 231)	BTXCo (N = 232)	Total (N = 463)
Inclusion criterion 3	1 (0.4); S009/065	-	1 (0.2)
Inclusion criterion 4	1 (0.4); S117/125	-	1 (0.2)
Inclusion criterion 5	1 (0.4); S009/065	-	1 (0.2)
Inclusion criterion 6	1 (0.4); S009/065	-	1 (0.2)
Inclusion criterion 7	2 (0.9); S009/065, S012/068	1 (0.4); S468/477	3 (0.6)
Inclusion criterion 8	1 (0.4); S009/065	-	1 (0.2)
Inclusion criterion 9/10	8 (3.5); S009/065, S012/068, S088/048, S218/014, S396/434, S401/439, S491/429, S497/1202	9 (3.9); S013/073, S089/047, S207/266, S286/298, S397/435, S398/436, S400/438, S402/440, S555/1168	17 (3.7)
Exclusion criterion 1	-	1 (0.4); S089/047	1 (0.2)
Exclusion criterion 7	-	1 (0.4); S038/008	1 (0.2)
No control visit conducted	-	4 (1.7); S094/052, S098/056, S243/272, S540/1185	4 (0.9)
Control visit time deviation (>7 days)	7 (3.0); S018/078, S020/269, S087/045, S147/199, S222/246, S322/360, S518/1139	5 (2.2); S006/058, S017/077, S223/247, S286/298, S587/1107	12 (2.6)
Forbidden/not stable medication	2 (0.9); S170/182, S497/1202	4 (1.7); S001/061, S004/064, S239/239, S549/483	6 (1.3)
Drug irregularities	2 (0.9); S032/001, S117/125	1 (0.4); S031/002	3 (0.6)
Only total amount of injected units known	-	1 (0.4); S129/205	1 (0.2)

^a Percentages based on the total number of patients in a patient group.
Data Source: Table 14.1.2.1.2.2, Table 16.2.1.1.2.1-5 and Table 2 of the SAP.
Abbreviations: N = number of patients.
Note that multiple reasons for discontinuation may have been reported for a single patient.

7.3.3.7. Baseline data

Overall, the groups were adequately matched. The duration of CD prior to study entry was 9.4plus or minus 7.7 (7.0) years in the NT 201 group and 9.2plus or minus 7.6 (7.0) in the Botox group (mean plus or minus SD; median).

The main type of cervical dystonia was specified as retrocollis in 113/463 patients (24.4%), and the two groups were not well matched for this feature: (retrocollis: NT 201, 71 subjects, 30.7%, versus Botox, 42 subjects, 18.1%). Other CD subtypes (anterocollis, laterocollis) were reasonably evenly distributed in the two groups.

7.3.3.8. Results for the primary efficacy outcome

Descriptive results for the primary efficacy variable are shown in the table below. In both treatment groups, a substantial fall in TWSTRS-Severity score was observed, and the results were numerically very similar in the two groups. The NT 201 group showed an improvement of 6.6 points from a baseline of 17.8 points, whereas the Botox group showed an improvement of 6.4 points from a baseline of 17.7 points. In both groups, the changes compared to baseline were statistically significant ($p < 0.001$ by ANCOVA), but this cannot be taken as solid evidence of efficacy because of the lack of a placebo control group.

The ANCOVA showed that the two treatments were equivalent within the pre-specified bounds. The mean improvement on NT 201 was numerically superior to that seen with Botox, which is expressed as a negative difference in the tables below.

The 95% confidence intervals included the possibilities that each drug was marginally superior to the other, as well as including zero, consistent with no difference. The intervals stretched further into negative differences (NT 201 superiority) than positive differences (Botox superiority), but basically showed the drugs to be very similar.

For a non-inferiority study, only the positive, upper bound of the 95% CI is relevant - if NT 201 is marginally superior to Botox, that does not represent a barrier to its registration, but it would potentially be a problem if NT 201 were inferior.

The ANCOVA put an upper limit on the superiority of Botox that ranged from 0.34 points to 0.51 points, depending on the precise method of analysis (full, final or simple ANCOVA) and the population considered (ITT or TPP). The upper 95% confidence limits of the differences between the two groups in the treatment-induced changes in the TWSTRS-Severity score were similar for the TPP (final model 0.38, full model 0.21, simple model 0.48) and the ITT populations (final model 0.40, full model 0.34, simple model 0.51).

Note that the upper limit of the 95% CI for the final TPP model (0.38) is considered to be the primary outcome of this study according to the prospective analysis plan, but the various models are reasonably concordant. By any method of analysis, the upper limit of the 95% CI is consistent with a trivial difference in treatment compared to the magnitude of the treatment effect seen with either drug. Thus, non-inferiority has been satisfactorily demonstrated.

The sponsor also presented a dose-related estimate of the 95% CIs for the change in TWSTRS-Severity score, which suggested that the two drugs showed broad equivalence across a range of doses. Note that this is not a dose-response curve. The dose administered was not randomised, but individualised for each patient, and thus acts partially as a marker for disease severity. Nonetheless, it is reassuring that the drugs were comparable across the range of doses assessed.

7.3.3.9. Results for other efficacy outcomes

All secondary endpoints were assessed in the TPP population rather than the ITT population. Given that the primary endpoint, discussed above, showed little difference in the ITT and TPP populations, this is acceptable.

Major secondary efficacy variables included TWSTRS-Severity at final visit, TWSTRS-Pain score at the control visit and final visit, and VAS pain scores at the control visit and final visit. For all of the measures listed, a statistically significant change was demonstrated by ANCOVA in comparisons versus baseline, for both drugs (but significance was not always demonstrated by the Wilcoxon test). This does not prove a true treatment benefit, because there was no placebo group and some of the observed improvement could be due to the placebo effect and regression to the mean. The magnitude of the apparent benefit was similar with both drugs, and there was no significant difference between the two treatments for any of the efficacy variables listed. By the Wilcoxon test, but not by the ANCOVA, there was a trend ($p = 0.06$) suggesting a between-treatment difference for final VAS pain minus baseline VAS pain, but pain scores had largely returned to baseline by this point. At the control visit, pain scores in both treatment groups were similar.

The Patient Evaluations of Global Response at the control visit and final visit, upon visual inspection of the distributions, suggests that patients responded in a similar fashion to both drugs. The ANCOVA did not detect any significant differences between the two treatments for this measure.

The time to subjective onset of treatment effect was very similar in the two treatment groups, and again ANCOVA did not detect any significant difference.

Onset of benefit, where noted, was typically 7 days after injection. Waning of effect was also similar with the two drugs, and typically occurred after 10 weeks.

The “duration of effect” was defined as the interval between treatments, and potentially includes some time waiting for the next injection after substantial loss of efficacy from the previous injection. On the other hand, subjects with cervical dystonia generally seek re-treatment before their problem reverts to the original, untreated state. The treatment interval is therefore likely to be an overestimate of the duration of a good effect, and an underestimate of the duration of at least some effect. The important issue in this context is whether the NT 201 formulation has a similar duration of effect as Botox, and that appears to be the case in the evidence from this study— both treatments had a mean duration of effect of approximately 95

days and a median of approximately 110 days. The two treatments were not significantly different by ANCOVA.

About 77% of each group were “responders”, defined as those with an improvement greater than or equal to 20% in their TWSTRS-Severity score at the control visit (NT 201 77.0%, Botox 77.3%); the minor difference was not significant.

The investigators Global Assessment of Efficacy showed very similar results with both treatments.

7.3.3.10. Subgroup analyses

The sponsor presented a range of subgroup analyses for this study, including those based on demographic characteristics, disease features and antibody status. No significant differences between treatments were observed (not shown).

One subgroup analysis was based on antibody status. Functional antibodies to botulinum toxin were measured with the mouse hemidiaphragm assay (HDA), which tests for the ability of the antibodies to block the weakening action of the toxin. At baseline, the HDA revealed positive antibodies in 18 patients (7.8%) in the NT 201 group and 25 patients (10.8%) in the Botox group. During treatment, two NT 201 recipients with a negative titre at baseline became positive (< 5 mU/mL), whereas in the Botox group four patients became HDA -positive (three patients < 5 mU/mL, one patient > 5 mU/mL). Four patients in each treatment group changed from HDA-positive to negative.

The efficacy results are displayed below for subjects with no antibodies, and those with low or high titres, at the control visit (CV) and final visit (FV). Higher titres appeared to be associated with reduced efficacy, but the numbers were small. There was no substantial difference between the two treatments within individual antibody subgroups. NT 201 is expected to be less immunogenic than Botox, and most of the antibodies detected are likely to be due to previous Botox treatment, so if HDA antibodies do cause reduced efficacy, this would be an argument in favour of using NT 201. The number of HDA-positive subjects in this study is too low to draw conclusions.

7.4. Spasticity

The sponsor performed two pivotal studies of NT 201 in the treatment of spasticity, both of which are described below. Supportive studies, including open-label extensions of the main studies.

7.4.1. Pivotal Study 0410

7.4.1.1. Study design, objectives, locations and dates

This Phase III pivotal study was titled “*Prospective, double-blind, placebo-controlled, randomized, multi-center trial with an open-label extension period to investigate the efficacy and safety of NT 201 in the treatment of post-stroke spasticity of the upper limb.*” It was performed in 23 centres from Poland, the Czech Republic and Hungary from June 2006 to January 2007, and also involved an extension phase described separately.

The objective was to investigate the efficacy and safety of individualised doses of NT 201 (170 U to 400 U) in comparison to placebo in the treatment of post-stroke spasticity, using a mixed population of botulinum-toxin naïve and pre-treated patients.

7.4.1.2. Inclusion and exclusion criteria

The main criterion for inclusion was that subjects had a diagnosis of post-stroke spasticity in a flexed wrist and clenched fist pattern, with both of the relevant muscle groups (wrist flexors and finger flexors) scoring greater than or equal to 2 in the Ashworth Spasticity Scale.

The inclusion and exclusion criteria in more detail, appear straightforward, and were aimed at recruiting a homogenous population without confounding factors or contraindications to treatment.

7.4.1.3. Study treatments

Subjects received a single treatment of NT 201 or matching placebo, at a dose selected by the investigator. The protocol specified a minimum dose of 170 U (3.4 mL trial medication) and a maximum of 400 U (8.0 mL). Subjects were then followed for a minimum of 13 weeks and a maximum of 21 weeks, before switching to an open-label extension study.

7.4.1.4. Efficacy variables and outcomes

Efficacy and eligibility assessments relied on the Ashworth scale, which is a simple, 5-point numerical scale rated by the investigator at each joint of interest:

- 0 = No increase in tone
- 1 = Slight increase in tone giving a “catch” when the limb was moved in flexion or extension
- 2 = More marked increase in tone, but limb easily flexed
- 3 = Considerable increase in tone - passive movements difficult
- 4 = Limb rigid in flexion or extension.

This scale has been used extensively in the literature on spasticity, and has been validated in the setting of post-stroke spasticity treatment by.¹⁶

The primary efficacy variable was the response rate at the Control Visit (Week 4), with a response defined as an improvement (reduction) of at least 1 point from the Baseline Visit (Day 0) in the Ashworth Scale score for wrist flexors. To be eligible at baseline, subjects had to score at least 2 points on the Ashworth scale at wrist flexors and finger flexors.

The Ashworth scale itself (rather than the response rate) was considered a secondary efficacy variable, as were the response rates in other muscle groups.

Additional efficacy variables included:

- Disability Assessment Scale (DAS), in one of four domains (hygiene, dressing, limb position or pain)
- Modified Carer Burden Scale (CBS), for subjects with a carer
- Investigator’s, Patient’s and Carer’s Global Assessment of Efficacy (GAE)
- Subjective time to onset of treatment effect
- Subjective time to waning of treatment effect

The DAS was rated as follows:

- 0 = No disability
- 1 = Mild disability (noticeable but does not interfere significantly with normal activities)
- 2 = Moderate disability (normal activities require increased effort or assistance)
- 3 = Severe disability (normal activities limited).

¹⁶ Brashear A, Gordon MF, Elovic E, et al. 2002. Intramuscular injection of botulinum toxin for the treatment of wrist and finger spasticity after a stroke. *N Engl J Med* 347(6):395-400.

The DAS can be applied to any of four domains (hygiene, dressing, limb position or pain). For each patient, the investigator was asked to identify one of these domains as the main target of treatment.

For the CBS, the patient's main carer was asked to assess the difficulty associated with each of the following tasks:

- A. Cleaning the palm of the affected hand
- B. Cutting the fingernails of the affected hand
- C. Cleaning the armpit of the affected arm
- D. Putting the affected arm through the sleeve (for example, coat, shirt, jacket)
- E. Applying a splint to the affected arm.

The rating scale used for the CBS was as follows:

- 0 = no difficulty
- 1 = a little difficulty
- 2 = moderate difficulty
- 3 = a great deal of difficulty
- 4 = cannot do the task

The GAE was based on a 4-point Likert scale:

- 1 = very good
- 2 = good
- 3 = moderate
- 4 = poor

7.4.1.5. Randomisation and blinding methods

Subjects were randomised with equal probability to active or placebo treatment, using a computerised randomisation scheme (RANCODE), stratified by centre. Blinding was maintained by using placebo vials that appeared identical to active vials.

7.4.1.6. Analysis populations

As described for other pivotal studies in this submission, the sponsor defined three populations: the Evaluable for Safety (EFS) population, the Intent-to-Treat (ITT) population, and the Treated-Per-Protocol (TPP) population. The ITT population was designated as the primary confirmatory analysis population for efficacy, but additional sensitivity analyses were performed with the TPP population.

7.4.1.7. Statistical methods

Analysis of the primary efficacy endpoint was based on a comparison of response rates at the Control Visit between NT 201 and placebo for wrist flexors in the ITT population, using a logistic regression model, and expressed as an odds ratios (OR). The primary analysis was performed without replacing missing observations (observed case analysis), but additional sensitivity analyses were performed with data imputation. The Ashworth Scale score at baseline, gender, age, body mass index (BMI), pre-treated versus naïve status, and pooled centre were entered into the model as covariates. Statistical testing was performed with a conventional two-sided alpha level of 0.05. Three models were derived: a full model with all pre-defined covariates, a final model adjusting for significant covariates (p less than or equal to 0.2), and a

simple model without adjustment for covariates. The full model was prospectively identified as the primary model for confirmatory testing.

Secondary efficacy variables were analysed in a descriptive and exploratory manner in the ITT and TPP populations, without adjustment for multiple endpoints.

7.4.1.8. Sample size

Sample size estimations were based on response rates to botulinum toxin in previously published studies of spasticity. The sponsor assumed that active treatment would be associated with a 3.5 fold higher chance (OR) of a response relative to placebo, and estimated that a chi-square test with a two-sided significance level of $\alpha = 0.05$ would have 90% power to detect an OR of 3.5, when the sample size in each treatment group was 63. Allowing for missing data, greater than or equal to 70 patients per treatment group were to be randomised, this target was exceeded.

7.4.1.9. Participant flow

Patient discontinuations were rare, but these are summarised below. The number of discontinuations is low enough that the study is unlikely to have been affected by withdrawal bias.

7.4.1.10. Major protocol violations/deviations

Major protocol deviations were rare, and the number of violations (NT 201 3 subjects, placebo 5 subjects) was acceptable for a study of this nature.

7.4.1.11. Baseline data

The groups were not well-matched in terms of the percentage of women with child-bearing potential, but the age and gender balance was acceptable, and there were no major demographic differences between the groups that were likely to have had an impact on the efficacy conclusions. The concomitant diseases were also reasonably balanced.

The patients' history in relation to their spasticity, the pattern of their spasticity and their previous treatments for spasticity are summarised in the tables below. The placebo group had a shorter duration of spasticity, on average, but the duration of the spasticity in each group was sufficiently long that subjects are likely to have had well-established and stable spasticity. A slightly higher proportion of patients in the active group had previously received botulinum toxin for spasticity. Overall, the minor differences noted are unlikely to have had any substantial impact on the efficacy results.

7.4.1.12. Results for the primary efficacy outcome

Results for the primary endpoint (response rates for Ashworth scores in wrist flexors): 68.5% of NT 201 recipients and 37.3% of placebo recipients were 1-point responders on the Ashworth scale. The OR in the full statistical model was 3.97, and this was highly significant ($p < 0.001$). Other models produced similar results

7.4.1.13. Results for other efficacy outcomes

When other muscle groups are considered, similar results are obtained.

The primary endpoint was based on a 1-point response, and with this definition several placebo recipients were classified as responders. Using a more restrictive definition of response, in which at least a 2-point improvement was required, showed that most subjects did not achieve a response of this magnitude, even with active treatment. Nonetheless, the odds ratio remained strongly favourable with this definition, because of the very low number of placebo responders (4.1%) compared to NT 201 responders (19.2%, $p = 0.007$).

The time to onset of treatment effect was approximately 4 days with active treatment, compared to 20 in the placebo group. Waning occurred at a median of 10 weeks with active

treatment. The time of waning was difficult to analyse in the placebo group because many subjects reported no onset of effect, and estimates varied from 5 to 10 weeks according to the method used.

The investigators GAE showed a clear difference in favour of NT 201, as did the patient and carer GAE.

For the Disability Assessment Scale (DAS), the sponsor differentiated between subjects who were 'sufficiently treated' and subjects who were only partially treated. According to the protocol, wrist and finger flexors required a treatment in every patient. Other muscle groups, such as elbow flexors, forearm pronators and thumb flexors, only required treatment if the corresponding clinical pattern was present and if the Ashworth score in the respective muscle groups was greater than or equal to 2. Some subjects had a clinical pattern for which the protocol required treatment of several muscle groups, but they only received treatment for some of those groups. These subjects were considered insufficiently treated. The DAS results were presented for all subjects, and for the 'sufficiently treated' subgroup.

For each subject, investigators were required to indicate the principal therapeutic domain being targeted. Improvements in the DAS results were noted in the active and placebo groups, but improvements were significantly greater with active treatment. Results for each individual domain (dressing, limb position, hygiene and pain) were also presented across the whole population, without regard to the principal therapeutic target of individual subjects, and these results also favoured active treatment.

The Carer Burden Scale (CBS) proved to be an insensitive measure, and did not show a significant difference between treatment groups.

7.4.1.14. Subgroup analyses

Subgroup analyses were presented with descriptive statistics. A broadly similar treatment effect was noted in men and women, in younger and older subjects, and in treatment-naïve and pre-treated subjects. For subjects with a baseline Ashworth score of 2 or 3, a clear increase in response rate was observed with active treatment. For the few subjects with an Ashworth score of 4, a response rate of 50% was observed in both the active and placebo groups. The lack of an apparent treatment effect in this group could reflect the very low patient numbers, but could also indicate that severe limb rigidity is less responsive to botulinum toxin. This would be the case, for instance, if subjects in this group had a degree of contracture (stiffness due to shortening of fibrous connective tissue elements, rather than muscle activity).

The response rate was lower in men who received NT 201 (55.6%) than in women (89.3%), a pattern noted in other submitted studies. This possibly reflects the higher muscle mass in men, leading to relative under-dosing. Numerical superiority over placebo was noted for both genders.

7.4.2. Pivotal Study 0607

7.4.2.1. Study design, objectives, locations and dates

This Phase III study was titled '*Prospective, randomized, observer-blind, parallel-group, multi-center trial to assess efficacy and safety of two different dilutions of NT 201 in patients with upper limb spasticity*'. It took place in 34 centres in Austria, France, Germany, Italy, Portugal, Spain, Switzerland, and the United Kingdom, from February 2007 to May 2008.

It differed from the previously described pivotal spasticity study in that the primary objective of the trial was to show non-inferiority of an injection of NT 201 high-volume dilution (20 U/mL) compared to NT 201 low-volume dilution (50 U/mL). It also recruited patients with a variety of causes for their spasticity, rather than just post-stroke spasticity.

Because this study did not have a non-NT 201 control group (placebo or competing botulinum toxin product), it should be considered supportive rather than pivotal in establishing the

efficacy of NT 201 in the treatment of spasticity. It sought to demonstrate that the NT 201 has a similar efficacy across a range of dilutions, which is necessary because the dilution of botulinum toxin is usually variable and individualised, depending on the site and indication.¹⁷

The sponsor explained the rationale for this study as follows:

'In previous studies, 2 mL per vial was used as the dilution volume (Study 0410). Therefore, for this study this dilution (50 U/mL) was considered as standard and was the low-volume dilution. As there are no data for NT 201 with a high-volume dilution, the purpose of this study was to show non-inferiority of the high-volume dilution with 5 mL per vial (20 U/mL) when compared with the standard low-volume dilution with 2 mL per vial (50 U/mL).'

7.4.2.2. Inclusion and exclusion criteria

Subjects were eligible if they were adults and had stable upper limb spasticity caused by stroke, brain injury, spinal cord injury, multiple sclerosis or cerebral palsy, regardless of whether they had been pre-treated or were treatment naïve. For multiple sclerosis patients, they had to have been relapse-free in the previous 6 months. The severity of the spasticity needed at baseline was the same as the previously described study (greater than or equal to 2 points on the Ashworth scale in wrist and finger flexors).

To allow assessment of the primary efficacy endpoint, the Disability Assessment Scale (DAS) at baseline had to be greater than or equal to 2 points for primary therapeutic target at both screening and baseline visits. (This scale was described for the previous study.)

For pre-treated subjects, the most recent injection session with botulinum toxin needed to be well-documented and there had to have been a sufficient therapeutic response for flexed wrist or flexed wrist and elbow. The total dose of the most recent injection had to have been not more than 400 U Botox (1600 units Dysport or 16.000 units Neurobloc) for the affected flexors of wrist or affected flexors of wrist and elbow. For flexor carpi ulnaris, the maximal dose was 50 units Botox (200 units Dysport or 2000 units Neurobloc) and for flexor carpi radialis, the maximal dose was 60 units BOTOX (240 units Dysport or 2400 units Neurobloc).

Females of childbearing potential had to have a negative blood pregnancy test.

Significant confounding illnesses or treatments were disallowed, as in previously described studies.

7.4.2.3. Study treatments

All subjects received NT 201 by intramuscular injections to the affected muscles. Dosing was individualised, and was based on the recommendations of the Worldwide Education and Awareness of Movement Disorders (WE MOVE) for each clinical pattern. The maximum dose was 400 units per subject, diluted as either 50 U/mL or 20 U/mL depending on the randomised treatment group.

Subjects received a single randomised treatment and were then followed for a minimum period of 12 weeks, with an optional follow-up of an additional 8 weeks. If a new injection of non-trial botulinum toxin was administered between Week 12 and Week 20, the final assessment was performed immediately prior to the second treatment.

¹⁷ In the proposed Product Information sheet, reconstitution with "an appropriate amount of solvent" is recommended. For two individual indications, the PI recommends specific dilutions: Glabellar frown lines, "Reconstitution: 50 units/1.25 mL (equivalent to 62.5 U/mL)"; Blepharospasm, "Initial recommended dose: 1.25 to 2.5 U (0.05-0.1 mL volume)/injection site" (equivalent to 25 U/mL). For Cervical Dystonia and Spasticity, the PI does not specify a dilution.

7.4.2.4. Efficacy variables and outcomes

The primary efficacy variable was the Disability Assessment Scale (DAS), as described for the previous study. For each subject, one DAS domain was designated as the principal therapeutic target. The primary efficacy endpoint was the DAS response rate at Week 4, defined as the proportion of subjects with an improvement of at least 1 point in the DAS for the primary therapeutic target from baseline.

Secondary endpoints were listed as follows:

- DAS response at Week 12 and Week 20;
- change in Frenchay Arm Test (FAT) from baseline to Week 4, Week 12 and Week 20;
- change in Ashworth scale for each treated flexor muscle group and for treated forearm pronators from baseline Week 4, Week 12 and Week 20;
- change in Activity of Daily Living (ADL) score (Barthel Index, selected items: feeding, grooming, toilet use, bathing, dressing) from baseline visit to Week 4, Week 12 and Week 20;
- change in investigator's and patient's Global Assessment of Treatment Response (GATR) at Week 4, Week 12 and Week 20;
- change in Passive Range of Motion (PROM) for wrist and elbow from baseline visit to Week 4, Week 12 and Week 20.

For the Frenchay Arm Test (FAT), the subject sits at a table with hands on lap, and is then asked to perform a number of standardised tasks:

- Stabilize a ruler while drawing a line with a pencil held in the other hand.
- Grasp a cylinder (12 mm diameter, 5 cm long) and set on its end, lift about 30 cm and replace it.
- Pick up a glass half-full of water.
- Remove and replace a sprung clothes peg from a 10 mm diameter dowel.
- Comb hair or imitate the motion required.

The total number of tasks completed gives the score.

Validation of this test was not discussed, which is a tolerable deficiency given that the FAT was only a minor efficacy variable.

7.4.2.5. Randomisation and blinding methods

Randomisation was performed with a software package, and blinding was maintained by using identical vials for active treatment and placebo.

7.4.2.6. Analysis populations

The sponsor defined three study populations, the Full Analysis Set (FAS), Per Protocol Set (PPS) and Safety Evaluation Set (SES), defined as in previous studies. Like the other non-inferiority studies in this submission, the primary efficacy analysis was performed on the PPS, with the FAS used for a secondary supportive analysis. This is acceptable.

7.4.2.7. Statistical methods

To analyse the primary endpoint, the sponsor calculated a two-sided 95% Newcombe-Wilson confidence interval (CI) for the difference between groups in the proportion of subjects with an improvement (reduction) of at least 1-point in DAS for the primary therapeutic target from baseline to Week 4. The lower bound of the Newcombe-Wilson CI for the difference of proportions was compared to the non-inferiority margin of -25%. An unadjusted Odds Ratio

(OR) was also estimated directly from the contingency tables, and an adjusted OR was calculated using a logistic regression model, with treatment and centre as factors and the DAS score at baseline as a covariate.

For the secondary efficacy endpoints, comparisons between treatments were presented with confidence intervals but these were considered descriptive and exploratory.

7.4.2.8. Sample size

To estimate the necessary sample size, the sponsor assumed that the response rate at Week 4 would be similar to that previously reported for Botox at Week 6, approximately 62.5%. Based on 10,000 simulations using the Newcombe-Wilson score method to construct CIs, it was estimated that, with 77 subjects in each treatment group, and an identical response rate of 62.5% in each group, the lower limit of the observed one-sided 97.5% CI for the difference in response rates between groups would be greater than -25%, and therefore lead to a non-inferiority conclusion, with 90% power.

To account for withdrawals and protocol violations, it was planned that 100 subjects per treatment group would be enrolled. This target was exceeded, and the final number of per-protocol subjects in each group was greater than 77 (20 U/mL n = 81; 50 U/mL, n = 84).

7.4.2.9. Participant flow

All randomised subjects received medication, and approximately 94% of these completed the study, which is acceptable for a study of this nature.

7.4.2.10. Major protocol violations/deviations

Major deviations included deviations in the injection procedure, missing or mistimed efficacy assessments, and violation of entry criteria and use of prohibited medication. Overall, the number of violations was acceptable.

7.4.2.11. Baseline data

The two treatment groups were reasonably well-matched at baseline in terms of demographics, but the proportion of males was higher in the 50 U/mL group. The characteristics of the subjects' spasticity were similar in the two groups, including aetiology and clinical pattern.

Importantly, the DAS domains designated as the primary therapeutic target were similarly distributed in the two treatment groups, and the baseline severity of the DAS was also broadly similar, allowing a meaningful comparison across groups.

7.4.2.12. Results for the primary efficacy outcome

The DAS response rate for the more concentrated NT 201 solution (50 U/mL) was numerically inferior to that seen with the more dilute preparation (20 U/mL), ; the difference in the PPS was 10.6% (95% CI -4.4% to 24.9%) in favour of the 20 U/mL group. The FAS analysis was very similar, as shown. The 95% CI (97.5% one-sided CI) surrounding this difference included the possibility of zero difference, and did not fall below the pre-specified non-equivalence limit of -25%. On this basis, the two dilutions should be considered equivalent in terms of efficacy. To determine whether the numerically better result in the 20 U/mL group represents random variation, or instead reflects that this dilution is preferable to 50 U/mL, additional studies would be required. There is no compelling reason for performing such studies, particularly given that the doses and dilutions for botulinum toxin are usually tailored for individual patients and conditions.

7.4.2.13. Results for other efficacy outcomes

Response rates at other time points also showed broad equivalence between the two treatment groups.

Response rates for the Ashworth scale in individual muscles generally showed minor superiority of the 20 U/mL dilution. With one exception (thumb flexors at Week 4), the between group differences were not significant, as indicated by 95% CIs for the differences that included zero.

The Frenchay Arm Test showed a low response rate overall. The 95% CIs for the differences between groups included zero at all-time points.

Patients' abilities to perform activities of daily living, as assessed by the Barthel Index, showed no overall difference between the two groups. The 95% CIs for between-group differences included zero for each time-point and activity.

The patients' and physicians' ratings on the Global Assessment of Treatment Response (GATR) are summarised below. The most common rating was 'moderate improvement', and the distribution of scores was similar in the two treatment groups.

The mean GATR scores were similar in the two groups, and the 95% CIs for the differences included the possibility of zero difference.

Finally, there was no significant difference between the two groups for Passive Range of Motion (PROM) scores.

7.4.2.14. Subgroup analyses

No subgroup analyses were planned or conducted in this study, which is appropriate considering it was a non-inferiority study only powered for the primary endpoint.

7.5. Supportive efficacy studies

7.5.1. Dose-ranging studies

7.5.1.1. Study 0527 (GFL)

7.5.1.1.1. Design

This Phase II study was titled: '*A prospective, randomized, double-blind, placebo-controlled, multicenter trial to determine the optimal dose of NT 201, free of complexing proteins, in the treatment of glabellar frown lines.*' It was a dose-ranging study limited to a single indication (Glabellar Frown Lines, GFL), and therefore only provides indirect evidence assisting dosing decisions for other indications. In addition to placebo, the doses tested were 10 U, 20 U and 30 U of NT 201, and subjects were randomised to each of these treatments with equal likelihood.

It was performed in 4 centres in the USA and Canada, from November 2006 to August 2007.

Subjects were eligible if they had moderate to severe GFL, with a severity of 2 or 3 on the Facial Wrinkle Scale (FWS) at maximum frown, were aged greater than or equal to 18 years, and had no major medical problems.

Subjects were injected in five sites during a single treatment procedure, as described for the pivotal GFL studies, and then followed for 180 days.

Response to treatment was assessed by the investigator, using the FWS. Patients also assessed their own facial wrinkling using a four-point scale, as described previously for Study 0724.

The main efficacy criterion was the Response Rate, where response was defined in two ways:

- the achievement of a 'none' (Grade 0) or 'mild' (Grade 1) score on the investigator's FWS,
- a 1-point improvement on the patient's self-assessment.

The following variables were considered co-primary:

- Response Rate at maximum frown on Day 30, by investigator's FWS

Response Rate at maximum frown on Day 30 by Patient's Assessment

Secondary variables consisted of the same Response Rates at Day 90. FWS ratings by an independent expert committee were considered tertiary and supportive.

The main statistical method used to compare doses was a Fisher's Exact Test comparing the Response Rates at Day 30 in each treatment group, with a two-sided significance level of $\alpha = 0.05$, adjusted for multiplicity by testing in a prespecified order. The primary analysis set consisted of all randomised patients with available FWS ratings at maximum frown on Days 0 and 30.

Secondary endpoints were analysed similarly, but the analyses were considered descriptive.

To ensure adequate sample size, the sponsor calculated the number of subjects needed to demonstrate significance in a Fisher's exact test for difference in proportions (with continuity correction) under the assumption of a two-sided alpha level of 5%, a power of 90%, a Response Rate in the placebo group of 25% and in the NT 201 group of 65%. These assumptions indicated that a sample size of at least 36 patients would be required per dose group, and a target of 45 per group (180 total) was set to allow for non-evaluable patients. In practice, 191 subjects were randomised, exceeding these targets.

7.5.1.1.2. Results

Completion rates were quite high for a study of this nature, and major protocol violations were relatively infrequent.

The patients were reasonably well-matched at baseline for demographics and previous use of botulinum toxin.

Analysis of the primary endpoint revealed a clear dose trend, and all active doses were significantly superior to placebo ($p < 0.0001$). Response Rates at Day 30, as assessed by the investigator according to the FWS, were 50.0%, 74.5%, and 91.7% in the 10 U, 20 U, and 30 U groups, respectively, compared to 2.1% in the placebo group.

For the Patient's Assessment at Day 30, Response Rates were 72.9%, 78.7%, and 89.6% in the 10 U, 20 U, and 30 U groups, respectively, compared to 19.1% in the placebo group.

These dose trends persisted to Day 90, though the actual Response Rates showed an expected decline from Day 30 to Day 90. Persistence of efficacy to Day 90 was better with the highest dose (30 U).

When an independent expert committee rated the patients' wrinkling from digital photos at Day 30, broadly similar results were obtained, though the dose trend was less clear and the 20 U group had a marginally higher Response Rate than the 30 U group. Given the fact that assessment via a photo is indirect, static and two-dimensional, this result should be considered less robust than results obtained from the blinded investigators.

When viewing the Response rates over the entire study period, it is clear that efficacy had been substantially lost by Day 120, and that higher doses (in the range tested) produced a better and longer-lasting response.

Overall, this study strongly suggests that 30 U of NT 201 is more effective at reducing wrinkling due to GFL than lower doses or placebo. Despite this, a dose of 20 U was chosen for the two pivotal GFL studies, presumably because the safety data showed a higher incidence of severe adverse events at the higher dose.

7.5.1.2. Study 9801 (CD)

7.5.1.2.1. Design

This study was a brief dose-ranging study in cervical dystonia, in which three doses of NT 201 were compared with a recommended dose of Botox. A single, blinded, randomised treatment was administered to the sternocleidomastoid (SCM) and splenius capitus (SPL) muscles of subjects with cervical dystonia (CD) and the decrease in dystonia was evaluated 2 weeks later with surface electromyography (EMG).

The study was conducted in 8 centres in Germany, from February 1999 to January 2000.

To be eligible, subjects had to have CD in a predominantly rotational form, with hypertrophy of SCM, that had not responded adequately to medication. Previous botulinum treatment was permitted, but not in the year prior to randomisation. Subjects were randomised to a low, medium or high-dose NT 201 group, or to Botox. The doses administered to the high-dose NT 201 group were the same as the doses administered to the Botox group. In each case, the SCM muscle received 1/3 of the total dose, administered to 3 sites 3cm apart, and the SPL muscle received 2/3, as follows:

- NT 201 10 U into SCM, 20 U into SPL
- NT 201 20 U into SCM, 40 U into SPL
- NT 201 30 U into SCM, 60 U into SPL
- Botox 30 U into SCM, 60 U into SPL

Subjects with an inadequate response at Visit 3 (Day 14), defined as a < 25% improvement in their TWSTRS Severity score, were eligible for a repeat dose of 60 U of study medication (NT 201 or Botox) into other affected muscles, but not to SCM, which was the muscle used to evaluate the efficacy of each treatment. Following the assessment at 14 days post-injection, subjects entered an uncontrolled follow-up period of 106 days.

The primary efficacy variable was the decrease in the maximum EMG motor potential (in mV) recorded over the injected SCM during maximal voluntary contraction. Note that this is a measure of voluntary muscle contraction, rather than a measure of dystonia, which is by definition involuntary, so it is not a measure of direct clinical relevance. Nonetheless, it is a useful surrogate marker of the efficacy of botulinum toxin in producing paralysis, which would be expected to translate into efficacy in treating dystonia, based on prior experience of using Botox for this condition. The use of a non-clinical measure means that, in some respects, this study could be considered a pharmacodynamic study, though it was submitted by the sponsor as a clinical dose-ranging study.

The reduction in SCM EMG potential was assessed using an ANCOVA model, with treatment as the main effect and baseline value as the covariate. The per-protocol analysis was considered primary.

Additional efficacy variables were the TWSTRS total severity score, TWSTRS pain subscale, and maximum diameter of SCM as assessed by ultrasonography on Day 14. The TWSTRS severity and pain subscale were analysed with an ANCOVA model, whereas the pain subscale was analysed categorically.

Sample size estimations were not based on empirical data, a significant limitation given that the study failed to achieve significant results and that the sponsor proposes that the lack of significant differences between groups is evidence of equivalence between groups. A sample size of 56 randomised patients (14 per group) was proposed to be able to detect a group difference of at least 1.1 standard deviations in the primary EMG parameter with a power of approximately 80%, assuming normally distributed EMG data. Whether this is a difference that is clinically meaningful is unknown. During the study, recruitment targets were lowered further,

to 10 subjects per group. Also, EMG data from one centre had to be excluded because of incorrect measurement techniques, so that the actual number of subjects per group in the per-protocol analysis ranged from 9-11, well below the 14 originally estimated. Thus, as a dose-ranging study, this study was under-powered and can only be considered weakly supportive of the idea that dosing is equivalent for NT 201 and Botox. (Other studies, including pivotal studies, addressed this question more directly, so this is not an important deficiency in the overall context of the submission.)

7.5.1.2.2. Results

Matching was not perfect, particularly with respect to gender, which reflects the low patient numbers.

All treatment groups showed a mean reduction in maximal EMG potential in SCM: -0.46 for the NT 201 10/20 U group, -0.51 for the NT 201 20/40 U group, and -0.58 for the NT 201 30/60 U group. Note that the apparent dose trend for mean changes was reversed for median changes. For the Botox group, a mean reduction of -0.67 was achieved. Statistical comparison of all four treatment groups showed no significant differences suggesting that, within the power limitations of this study, the doses and treatments were equivalent.

Secondary efficacy variables did not show a dose trend. Reductions in TWSTRS total severity scores were -2.7, -1.5, and -5.2 for the low, medium and high-dose NT 201 groups, respectively, and -2.8 for the Botox group. For this measure, all three NT 201 groups were broadly similar to Botox, with no significant differences by ANCOVA. The high-dose and medium-dose NT 201 groups were significantly different from each other ($p = 0.013$), but this result should be interpreted with caution given that there was no consistent dose trend.

Similarly, the reductions in maximum diameter of SCM estimated by ultrasound were similar across the four treatment groups, with reductions of -3.0, -2.1, and -2.2 for the low, medium and high-dose NT 201 groups, and -2.3 for the Botox group.

TWSTRS pain subscales did not change significantly in any treatment group and no differences between groups were noted for this parameter. Basically, scores were low and remained low following treatment.

7.5.2. Other efficacy studies

7.5.2.1. Study 0520 (GFL)

7.5.2.1.1. Design

This study was a Phase III study, titled '*A prospective, randomized, double-blind, placebo-controlled, multicenter trial with an open-label extension period to investigate the efficacy and safety of NT 201, free of complexing proteins, in the treatment of glabellar frown lines.*' The Main Period was performed in 10 centres in Germany, from October 2006 to July 2007, followed by an Extension Period from February 2007 to November 2007. The Extension Period is described separately.

The design was similar to the two pivotal GFL studies. Eligible subjects were adults greater than or equal to 18 years, with moderate to severe GFL, scoring 2 or 3 on the Facial Wrinkling Scale (FWS) at maximal frown, in a stable medical condition. Additional entry criteria were very similar to those in the pivotal GFL studies, and were largely aimed at avoiding confounding conditions or contraindications to treatment.

Subjects were randomised at a 2:1 ratio to NT 201 (20 U in 0.5 mL) or matching placebo. Treatment was administered intramuscularly in a single session to 5 sites, as in the pivotal GFL studies. Subjects were followed for 120 days in the Main Period, and then entered an OLEX Period in which they received another dose and were followed for another 120 days.

Treatment was assessed with the FWS and a Patient's Assessment of change in frown lines, on a 9-point scale ranging from - 4 to + 4, as defined for the pivotal studies.

There were two co-primary endpoints. The first was the percentage of responders at maximum frown on Day 30 of the main period, where a responder was defined as a subject scoring 0 ('none') or 1 ('mild') on the FWS, as rated by the investigator. (Note that this is a simpler endpoint than the composite treatment success defined for the pivotal studies.) The second co-primary endpoint was the percentage of responders at maximum frown on Day 30, with response defined as a score of at least +2 on the Patient's Assessment.

Secondary endpoints included the percentage of responders at other time points, using the FWS and Patient's Assessment, and the percentage of responders during the OLEX period. The percentage of responders at rest and the time to subjective onset of effect were considered tertiary endpoints.

The primary endpoints were analysed with the Cochran-Mantel-Haenszel (CMH) method, with centre as a stratification variable, and a one-sided significance level of 0.0125). The Full Analysis Set (FAS) was considered the primary population for analysis, and missing values were imputed with a LOCF approach. Analysis of other variables was considered descriptive.

7.5.2.1.2. Results

All randomised subjects were treated and entered the Full Analysis Set. A total of 31 had major protocol deviations. The most common violations were mistimed visits and use of forbidden medication. Overall, the number of violations was acceptable for a study of this nature.

There were no important mismatches between the demographic characteristics of the active and placebo groups.

The groups were also reasonably matched at baseline in terms of the FWS scores at maximal frown. There was, however, a slight excess of placebo recipients with 'severe' FWS ratings (86%) compared to the active group (80%). This slight mismatch would be expected to bias the study in favour of active treatment, by making it less likely for placebo recipients to achieve a rating of 'mild' or 'none'. In practice, this is unlikely to have been important because of the marked difference in outcome in the two groups.

The proportion of responders at maximal frown, by the investigator's FWS, was 52.5% in the active group and 0% with placebo. This was highly significant (one-sided $p < 0.0001$) by the CMH method. By the Patient's Assessment, the proportion of responders in the active group was higher (67.5%), and a single placebo recipient self-rated as a responder (1.1%). This difference was also highly significant by the CMH method (one-sided $p < 0.0001$).

Responder rates at other time points were also strongly in favour of active treatment, with no placebo responders observed for the FWS rating, and very few placebo responders according to the Patient's Assessment. The highest response rates in the active group were observed at Day 30, followed by a gradual decline. By Day 120, the response rate in the active group was 9.5% by the investigators FWS and 21.3% by the Patient's Assessment.

In general, the response rates at rest were lower, as might be expected for a paralytic agent, but there was still a clear difference between the active and placebo groups.

The time to onset of treatment effect, judged subjectively by patients, was broadly consistent with other studies of NT 201 in GFL. The most common day on which an effect was noted was Day 3, two days post-injection. The majority of subjects receiving NT 201 had recorded the onset of an effect by Day 5.

7.5.2.2. Study 3002 (GFL)

7.5.2.2.1. Design

This study was Phase III non-inferiority study comparing NT 201 and standard botulinum toxin type A (Vistabel) in the treatment of GFL. It was titled '*A prospective, multicenter, randomized, rater- and subject-blind, parallel group trial to investigate the non-inferiority of NT 201, free of complexing proteins, in comparison with Clostridium botulinum toxin type A in the treatment of glabellar frown lines*, and was conducted in 20 centres in Germany and Austria from November 2008 to May 2009.

Because the study did not employ a double-blind design, it cannot be considered pivotal, though in other respects it was adequately designed.

Eligible subjects were females aged 18-50, with moderate to severe GFL, scoring 2 or 3 on the FWS. Other entry criteria were essentially the same as previously described GFL studies.

Subjects were randomised at a 3:1 ratio to NT 201 24 U (in 0.6 mLs) or Vistabel 24 U (in 0.6 mLs). They received a single intramuscular treatment, as described for other GFL studies, and were followed for 12 weeks.

The primary endpoint was the Response Rate, defined as an improvement of greater than or equal to 1 point in the FWS score at maximum frown on Day 28, based on an independent blinded rater panel who viewed digital photos. Note that blinding the rater does not necessarily lead to an unbiased outcome, because the person taking the photograph is likely to have had some scope for modifying the appearance of the facial wrinkles.

Secondary endpoints included the Response Rate at maximal frown on Day 84, Response Rates at rest, and Response Rates based on the unblinded site investigator's assessment of the FWS. Patients also performed a self-assessment at maximal frown on a 4-point scale, as well as a Patient's Global Assessment (PGA) of change on a 9-point scale ranging from - 4 to + 4, as described previously.

The primary endpoint was analysed with frequency tables, applied to the Per-Protocol Set (PPS). For the primary efficacy parameter, defined as the difference in response rates of the two treatment groups, a two-sided 95% Newcombe-Wilson confidence interval (CI) was calculated. Inferiority of NT 201 was to be inferred if the 95% CI extended below the non-inferiority margin of -0.15. To check for robustness of the results, the analysis was to be repeated in the Full Analysis Set (FAS) with a variety of imputation methods including baseline values carried forwards, next-observation-carried-backwards, and an observed case approach. Statistical analyses performed on secondary endpoints were considered descriptive and exploratory.

Sample size estimates were based on 10,000 computer simulations, using the Newcombe-Wilson score method to construct the 95% CI. The sponsor estimated that 240 NT 201 subjects and 80 Vistabel subjects (total 320) would produce a lower limit of the observed two-sided 95% CI for the difference in response rates greater than -0.15 with 90% power when the expected response rate for both treatment groups was 80% at Day 28. The recruitment target was increased to allow for protocol violations and withdrawals. Response rates were considerably better than the sample size calculations anticipated, but the overall sample size was appropriate. In practice, the actual 95% CI was narrow and easily exceeded the lower limit of the non-inferiority threshold.

7.5.2.2.2. Results

Most subjects completed the study. Overall, the number of withdrawals and violations was acceptable for a study of this nature.

Baseline characteristics, including FWS scores, were reasonably well matched across the treatment groups.

The response rate on Day 28 with respect to the FWS at maximum frown was 96.4% in the NT 201 group and 95.7% in the Vistabel group. The 95% CI for the difference in response rates between treatment groups (-3.2%; 7.1%) included zero and the lowest extent of the 95% CI was only -3.2%, well above the pre-defined non-inferiority margin of -15%. The results imply that NT 201 is similar in efficacy to Vistabel and, at worst, is associated with a response rate only 3% worse than Vistabel. Sensitivity analyses based on the FAS produced similar results.

Response Rates based on the unblinded assessment of site investigators were broadly concordant with the primary results: 98.9% of the NT 201 group and 95.7% of the Vistabel group showed an FWS response at 4 weeks.

Consistent with the primary endpoint, there was no apparent difference between the two treatments based on assessments performed on Day 84, assessments at rest, or on the subjects' self-rating of appearance and the PGA.

On balance, the results were consistent with equivalence of NT 201 and Vistabel, but firm conclusions cannot be drawn from this study because of the partially unblinded design.

7.5.3. Repeat-dose studies and persistence of efficacy

7.5.3.1. Study 0609 (GFL)

7.5.3.1.1. Design

This Phase III extension study was titled '*A prospective, open-label, multicenter, repeat-dose trial to investigate the safety and efficacy of NT 201, free of complexing proteins, in the treatment of glabellar frown lines.*' Because it was open-label and lacked a control group, it provides only weak support for the sponsor's efficacy claims. It does allow some assessment of the persistence of efficacy of NT 201, but this is not a robust assessment because of the unblinded nature of the treatment and the potential for withdrawal bias; only patients satisfied with the treatment would be expected to remain in the study.

The study was performed in 26 centres in Canada, Germany, and the USA from June 07 to December 09.

Patients with GFL were eligible if they completed one of the feed-in studies: MRZ 60201-0520/1, MRZ 60201 0527/1, MRZ 60201-0724/1, or MRZ 60201-0741/1. At the time of submission, only patients from the first two of these studies had the possibility of one year's follow-up. Both of these studies were non-pivotal GFL studies that are described separately in this report (Study 0520, and Study 0527). The dose in the feed-in studies was variable, consisting of placebo, NT 201 10 U, 20 U or 30 U.

Subjects received up to 20 U NT 201 on their first visit in the extension study. Re-injections could be performed after 85 days, with injection considered Day 0 of the next cycle, and treatment continued for up to 8 cycles.

The key efficacy parameter was the percentage of responders at maximum frown, where a responder was defined as a subject scoring 'none' or 'mild' according to FWS (described for the pivotal GFL studies), as assessed by the investigator at each cycle. The results were summarised with descriptive statistics.

7.5.3.1.2. Results

Patient disposition is summarised below. About 10% of subjects discontinued prematurely, which might be expected to introduce some withdrawal bias. The most common major protocol violations consisted of visits outside the specified time window (9.4% of subjects) and use of forbidden concomitant medications (7.7%).

The response rate observed at each cycle is shown in the figure below. There is a minor improvement in response rate over time, which could reflect withdrawal of poorly responsive patients. The response rate according to the dose administered in the relevant feed-in study

showed no overall difference for the different dose groups once they entered the open-label extension.

As can be seen, the number of patients contributing data to later cycles was low, which largely reflects the fact that the study is on-going and not all patients have been followed for long enough. Given the open-label, uncontrolled nature of this study, no firm conclusions can be drawn from these results, but they do not indicate a substantial waning of efficacy with continued use.

Other endpoints, such as the response rate based on the subject's FWS rating, or on the FWS assessed at rest instead of at maximum frown, showed a similar persistence of efficacy over the course of the study (not shown).

7.5.3.2. Study 0408/2 (CD)

7.5.3.2.1. Design

This study was a double-blind extension of the pivotal Study 0408. Subjects were eligible if they finished Study 0408 and required re-treatment. The dose in the Extension Period was 240 U or 120 U of NT 201, based on the original randomised treatment, reconstituted in 4.8 mL 0.9% NaCl and continued for up to 5 injection sessions. Dose reduction was allowed in case of safety concerns. The timing of repeat doses was based on clinical need, but could not be less than 6 weeks.

The primary efficacy variable was the mean change in the TWSTRS-Total Score, as previously defined for the Main Period of the pivotal study. This is the same efficacy variable used in the pivotal study, and is therefore appropriate for assessing the persistence of efficacy. For this variable, comparisons between the treatment groups were performed using descriptive two-sample t-tests for the mean difference in the changes from previous injection session.

Comparisons within each treatment group used a descriptive one-sample t-test and were based on a comparison of the pre- and post-treatment TWSTRS score (from each Injection Visit to the next Control Visit, which was 4 weeks later). Note that the treatment effect inferred from such a comparison includes any placebo response, so the finding of a significant p-value does not provide robust evidence of efficacy.

Other TWSTRS subscores and the Patient Evaluation of Global Response (PEGR) were considered secondary efficacy variables.

7.5.3.2.2. Results

In general, efficacy was maintained throughout the study, as indicated by a return to similar post-treatment scores following each treatment. Scores for the higher dose (240 U) were usually slightly better than those observed for the lower dose (120 U), but the between-group differences were small and not statistically significant.

A similar analysis was performed on the individual domains of the TWSTRS score, and this showed a similar preservation of therapeutic effect over the course of the Extension Period.

The PEGR showed that a mean score of 1.5 to 2.3 across the different groups and cycles, consistent with a moderate improvement overall (+ 1 = slight improvement, + 2 = moderate improvement, + 3 = marked improvement). There was no major change in this parameter over the course of the study. Statistically significant between-group differences (p less than or equal to 0.036) were observed for the 2nd and 3rd Injection Interval, in favour of the 240 U group.

7.5.4. Study 0605 (CD)

7.5.4.1.1. Design

This Phase IV study was titled '*Prospective, single-arm, multicenter study to investigate the efficacy and safety of NT 201 (botulinum neurotoxin type A free of complexing proteins) and the*

duration of treatment effect after one injection session and in long-term treatment in subjects with cervical dystonia'. It was conducted in 17 centres in Germany and was completed in 2010.

Because it lacked a control group, it is impossible to distinguish placebo effects from true efficacy, and it therefore only provides weak support of the sponsor's efficacy claims.

Subjects with cervical dystonia (CD) received a single, open-label injection of NT 201 in the Main Period of the study, then returned for up to four additional injection sessions during the Long-term Extension Period. The dose per injection session was individualised according to standard clinical criteria outlined in the protocol, and ranged from 50 to 300 U. They were followed for 4 weeks post-injection in the Main Period, but for up to 121 weeks if they entered the Extension Period.

The entry criteria were similar to the pivotal studies in CD: subjects were treatment-naïve or pre-treated, aged greater than or equal to 18 to < 76 years, and with a clinical diagnosis of CD with a need for injection. Their TWSTRS total score at baseline had to be greater than or equal to 25, with TWSTRS severity score greater than or equal to 10 and disability score greater than or equal to 3. Pre-treated subjects had to have adequate documentation of the previous two injection sessions,

a stable response, and had to have received a dose of less than or equal to 300 units of Botox or Xeomin, or less than or equal to 1,200 units of Dysport in the most recent injection session. At least 10 weeks had to have passed between the most recent injection session and the Baseline Visit of this study.

The primary efficacy variable was the change in TWSTRS total score from baseline to Week 4.

For this parameter, a 95% CI was calculated for the Full Analysis Set (FAS) using a variety of methods of handling missing data (LOCF, median imputation and observed cases); for the Per Protocol Set (PPS), the analysis was done using observed cases. An analysis of covariance (ANCOVA) was also performed, using the change from baseline in TWSTRS total score as dependent variable and baseline TWSTRS total score, gender, age, centre and pre-treatment status (naïve versus pre-treated) as covariates.

Changes in TWSTRS subscores were considered secondary variables, as was a Global Assessment of Efficacy by Investigator (GAEI, a 4-point Likert scale) and Patient Evaluation of Global Response (PEGR, a 9-point Likert scale ranging from -4 to +4, as described for the pivotal CD studies). Subjects also completed a study diary, which included a measure of discomfort called the Dystonia Discomfort Scale (DDS). For this scale, the subject was asked to record discomfort on a daily basis, before bedtime. The scale was divided into 5% steps and ranged from 0% (lack of any complaints) to 100% (maximal discomfort); it was stressed to subjects that 100% reflected the untreated condition, so pre-treated subjects would be expected to start with < 100% discomfort. Subjective times to onset and waning of effect were also noted, as well as the 'duration of treatment effect' defined as the time from injection to the time when the patient indicated a need for re-injection.

The sponsor indicated that the analysis of the primary efficacy variable was considered confirmatory, and secondary efficacy variables were considered descriptive. In the absence of a control group, though, none of the changes in any of the efficacy parameters can be interpreted as a reliable indicator of efficacy.

7.5.4.1.2. Results

Baseline demographic and disease characteristics are shown in the table below. The population studied appears broadly representative of the intended target population.

The primary endpoint, TWSTRS total score, showed an improvement from a mean baseline value of 39 to a mean of 27.4 at 4 weeks (FAS, LOCF); the difference was estimated to be 11.7 points (with a minor discrepancy in the total attributable to rounding errors). The 95% CIs

indicate that the improvement was statistically significant, and sensitivity analyses (using other methods of imputation or using the PPS population) confirmed that the observation was robust. It is unclear, however, how much of this improvement represents a placebo response. (In the pivotal placebo-controlled CD study, Study 0408, the improvement of the TWSTRS total score in the placebo group was approximately 8 points.)

Over subsequent cycles, the improvement observed within each injection cycle (from the pre-injection TWSTRS total score of each cycle to the score 4 weeks later) progressively decreased in magnitude. The pattern was not suggestive of a progressive loss of efficacy, because the TWSTRS total scores showed continued mild improvement over the course of the study, compared to the original study baseline. Instead, the reduction in the changes within each cycle could indicate some lingering benefit from previous cycles prior to each injection. There could also have been drift in the application of the TWSTRS scores, which is necessarily somewhat subjective. It is impossible to draw any firm conclusions without a control group.

Similar changes were observed in the component subscores of the TWSTRS, with a progressive fall in the amplitude of the changes per cycle, but an overall improvement in the scores across multiple cycles (not shown).

The GAEI and the PEGR showed that most investigators and patients rated the treatment as effective (the most common responses were 'Good' on the GAEI and '+ 3' or 'Marked improvement' on the PEGR). These ratings did not deteriorate with repeat dosing. The GAEI results are tabulated below, and the PEGR in the subsequent table.

The subjective time course of the treatment effect was fairly consistent across multiple cycles, with an onset 9-11 days after injection, waning beginning from 6-7 weeks, and the duration of effect estimated at 9-10 weeks.

The Dystonia Discomfort Scale (DDS) showed subjective improvements in discomfort with each treatment. In the absence of a placebo control group, it is unclear to what extent this represents a true therapeutic effect. There was, at least, no apparent waning of efficacy over the course of 5 cycles. (Note that, in the table below, even-numbered visits represent pre-injection baseline values, and odd-numbered visits are 4 weeks post-injection.)

7.5.4.2. Study 0433/2 (BLEPH)

7.5.4.2.1. Design

This was an open-label extension study in which patients with blepharospasm, who had completed the pivotal placebo-controlled Study 0433 and required further treatment, received unblinded NT 201 for up to 5 sequential treatments. To be eligible, they had to have a JRS Severity subscore greater than or equal to 2 points at the baseline of the Open-Label Extension (OLEX) Period.

All treatment in the OLEX period was with intramuscular NT 201, up to 50 U per eye, with dose adjustment permitted based on the response to previous injections.

Endpoints in the OLEX phase were tertiary. They included changes in the JRS Severity subscore, the JRS Frequency subscore, and the JRS Sum score, relative to baseline and to the previous injection cycle. They also included changes in the BSDI, the PEGR and GAEI, as defined for the pivotal study.

Statistical analyses in the OLEX period were considered descriptive and exploratory, which is appropriate considering its unblinded and uncontrolled design. For changes in JRS scores and BSDI, one sample t-tests were performed, with no replacement of missing data.

7.5.4.2.2. Results

Changes in the JRS Severity subscore, JRS Frequency subscore and JRS sum score are summarised below for each cycle. Improvements in the first injection cycle were generally

larger than later improvements, but all within-cycle changes showed clear statistical significance. Interpretation of these observations is difficult without a control group. A reduction in the amplitude of the within-cycle changes does not necessarily indicate a waning of efficacy, because subjects could still be benefiting from improvements achieved with earlier injections. In fact, visual inspection of the JRS-Severity and JRS-Frequency subscores over the course of the OLEX period suggests a gradual improvement in both subscores with repeated treatment (see figures below the table).

Changes in the BSDI showed a similar pattern, with improvements following each injection as well as an overall improvement through the course of the study. In the absence of a control group, it is unclear how much of the observed improvement is due to the placebo effect.

The PEGR results showed that most subjects reported improvement, and 'Marked improvement' was the most common response over the course of the OLEX period. There was no substantial shift in the distribution of PEGR ratings over time, suggesting a consistent subjective response to treatment.

Similarly, most of the physician ratings (GAEI) were positive, and showed persistence or even improvement in efficacy over the course of the OLEX period (see the table below). It is unclear to what extent this represents a true therapeutic effect, because there was no control group, but the investigators did not seem to notice any substantial waning of efficacy with continued use.

The duration of the treatment effect in this study, defined as the period between sequential injections, was consistent with other studies, and ranged from about 10-12 weeks. Note that some other repeat-dose studies defined treatment duration as the time to the patient's request for retreatment. By adding the logistical delay involved in organising the treatment after the patient has requested it, the definition used in this study is likely to have inflated the duration of treatment. On the other hand, some residual benefit from previous injections is likely to have been present at the time of retreatment, as suggested by visual inspection of the JRS scores (see figures above).

7.5.4.3. Study 0410/2 (SP)

7.5.4.3.1. Design

This study was an open-labelled extension of Study 0410 (described previously), in which subjects completing the Main Period of Study 0410 remained under observation and received unblinded treatment with NT 201 for an additional 5 injection sessions. The OLEX period had no specific entry criteria as it was expected that all subjects from the Main Period would enter the OLEX period.

All subjects in the OLEX period received NT 201, with the protocol specifying doses up to a maximum of 400 U (8 mLs of 50 U/mL); the maximum dose actually administered was 500 U. Doses were individualised, but were based on standard recommendations for the treatment of post-stroke spasticity, in particular the guidelines produced by WE MOVE, which include default doses for individual muscles as well as dose modifiers for different clinical situations, such as the patient's muscle bulk and the response to previous treatments.

The OLEX period had a minimum duration of 48 weeks and a maximum of 69 weeks.

Efficacy analyses in the OLEX were considered secondary, but the primary variable from the Main Period could be considered the most important variable for the OLEX Period. This was the response rate at Week 4 post-injection, with a response defined as an improvement of at least 1 point from baseline in the Ashworth Scale score for wrist flexors. Additional efficacy variables included the Ashworth scale itself (rather than the Ashworth response rate), as well as results in other muscle groups, the Disability Assessment Score (DAS), and Investigator's, Patient's and Carer's Global Assessment of Efficacy.

7.5.4.3.2. Results

The Response Rates for different muscle groups are shown in the table below, with the results in Wrist Flexors considered most important. For most muscle groups, a broadly consistent response was seen across the OLEX period, particularly in the first four cycles. (Note that the number of subjects in the 5th injection cycle was low.) By the Wilcoxon signed rank test, changes to the Ashworth scores at the wrist were statistically significant for every cycle, but this cannot be considered robust evidence of efficacy because there was no placebo group and all response rates include the placebo response. (In the pivotal, placebo-controlled phase of Study 0410, 37.3% of placebo recipients were responders).

Changes in the DAS Score were generally favourable at each treatment cycle. In the table below, changes in DAS for the principal therapeutic target are shown for control visits versus the previous injection visit (upper part of the table) and for injection visits versus previous injection visits (lower part of the table). Significant improvements occurred within each treatment cycle, but little change occurred from one cycle to the next.

The Global Assessments of Efficacy for investigators, patients and carers are shown graphically below. In most cases, treatment response was assessed as 'Good' and this did not change substantially over the course of the OLEX period.

Overall, this study was reasonably reassuring, suggesting that the efficacy of NT 201 is maintained over multiple cycles when treating post-stroke spasticity. Firm conclusions about the magnitude of the therapeutic effect cannot be drawn, however, because treatment was unblinded and lacked a control.

7.5.4.4. Study 0520/2 (GFL)

Study 0520 has been described previously. It was a supportive study investigating the efficacy of NT 201 in the treatment of GFL, in comparison to placebo. Like many other submitted studies, it included an open-label extension (OLEX) period, the results of which are included here. Only a single dose was provided in the OLEX phase, so these results add little to the overall multi-dose data.

On completing the blinded phase of the study, subjects who still required treatment entered an OLEX period and received unblinded NT 201 at a dose of 20 U in 0.5 mL.

Analyses in the OLEX period were considered descriptive and exploratory. The main efficacy parameters of interest were the percentage of responders at maximum frown on Day 30, according to two definitions of response: a subject scoring 0 ('none') or 1 ('mild') on the FWS, as rated by the investigator; and those with a score of at least + 2 on the Patient's Assessment.

The Response Rates are shown in the tables and figure below. Overall, subjects who had received NT 201 in the Main Period and thus received their second dose on the OLEX Period had a slightly better Response Rate than subjects who had received placebo in the Main Period, particularly in the first 30 days. This is suggestive of a cumulative effect with multiple treatments, but the results are not conclusive because formal statistical comparisons were not performed and the study was not powered for such a comparison. There was, at least, no evidence of waning efficacy for subjects receiving their second dose.

7.6. Analyses performed across trials

7.6.1. Pooled analysis of Studies 0724 and 0741 (GFL)

Studies 0724 and 0741 had almost identical designs, and the sponsor presented a pooled analysis of both studies. This analysis did not add new insights compared to the original studies. Across the pooled placebo group, there was not one responder for the primary endpoint (CETS).

7.6.2. Comparisons with other botulinum neurotoxin products

The sponsor grouped all active-controlled efficacy data, as shown in the tables below. This exercise highlighted the overall consistency of the results of the individual studies, which showed equivalence between NT 201 and Botox. Note that Study 0607 compared two different dilutions of NT 201.

7.6.3. Comparisons across studies with the same indication

Results in studies assessing glabellar frown lines were pooled, as described above. For studies assessing blepharospasm and spasticity, the major studies had some methodological differences, preventing pooling, but the major efficacy variables were similar. Assessing the blepharospasm studies side-by-side, as in the table below, suggests that the results are broadly consistent. Results in the spasticity studies are shown in the figure below, as reflected in Ashworth response rates, and again the effect of NT 201 across studies was broadly consistent.

For cervical dystonia, the primary endpoints differed across studies, making comparison more difficult. The primary result in Study 0408 was an improvement of approximately 10 points from a baseline of approximately 42 points in the TWSTRS-Total score (with slightly different results in the 240 U and 120 U dose groups), and the primary result in Study 0013 was an improvement of 6.6 points from the TWSTRS Severity subscore, from a baseline of 17.8, following a dose of 70-300 U.

7.6.4. Comparisons across different indications

A formal comparison of efficacy across different indications is not possible, but the common mechanism of action in each indication is weakening of muscle. That means that efficacy in each indication is supportive of the general claim that NT 201 is equivalent in efficacy to other botulinum toxins, and the set of submitted studies are mutually supportive even when the indications differ.

The sponsor attempted to show consistency of effect across indications by expressing the improvement in the primary efficacy criterion as a percentage. The results of this informal exercise, shown below, are generally reassuring. The subsequent figure shows that investigators were also consistent in their positive rating of the response to NT 201 across the indications tested. In repeat-dose studies, investigator ratings remained positive across multiple dose cycles, regardless of the indication.

The time course of the response differed across indications, but was broadly consistent with the shared mechanism of action and previous experience with other botulinum toxins. NT 201 begins to work in a few days, starts to wane in efficacy after 7-10 weeks, and usually requires repeat treatment in approximately 80-100 days.

7.7. Dosing considerations

In all of the major studies, dosing for treatment-naïve patients was based on treatment guidelines for Botox, but was generally individualised for each patient. (For glabellar frown lines, dosing was fixed at as a single intramuscular dose of 20 U). For patients who had already received Botox treatment, the dose of NT 201 was matched to the previously stabilised Botox dose. This approach produced clear efficacy in the placebo-controlled studies, and showed that NT 201 and Botox were equivalent in the active-controlled studies. The pharmacodynamic studies were also consistent with dose-equivalence between Botox and NT 201, though these studies were small.

It is expected that clinicians using NT 201 will adjust the dose according to baseline factors such as indication, muscle mass and severity of symptoms, response, and then adjust subsequent doses according to response and complications, just as they already do for Botox and other botulinum toxin preparations. Because of the 1:1 dose equivalence of NT 201 and Botox, dosing

guidelines for Botox could be adapted directly for NT 201; conversion of doses for Dysport recipients will require the same equivalence-ratio as is already used for Botox: Dysport conversions.

The proposed PI gives adequate instructions to guide dosing, as summarised previously.

The tables below indicate the mean doses actually used in the major studies, where these were not strictly limited by the protocol.

7.8. Blepharospasm

7.8.1. Pivotal Study 0433

7.8.1.1. Study design, objectives, locations and dates

This Phase III pivotal study was titled: '*Prospective, double-blind, placebo-controlled, randomized, multicenter trial with an open-label extension period to investigate the efficacy and safety of NT 201 in the treatment of blepharospasm.*' It was conducted in 34 centres in the USA and Canada, from October 2006 to May 2008.

The objective of the study was to investigate the safety and efficacy of NT 201 in comparison to placebo, when used to treat subjects with blepharospasm who had already been treated with Botox.

7.8.1.2. Inclusion and exclusion criteria

Subjects were eligible if they were male or female outpatients aged 18 to 80 years, with a clinical diagnosis of benign essential blepharospasm (BEB), previously treated with Botox, and with a Jankovic Rating Scale (JRS) Severity subscore greater than or equal to 2 at baseline. (This scale is discussed below).

Overall, the inclusion and exclusion criteria were reasonable, and were aimed at recruiting a homogenous population of subjects without confounding factors or significant contraindications to treatment. The requirement that subjects have a JRS Severity subscore greater than or equal to 2 at baseline is consistent with standard recommendations for botulinum toxin treatment. This ensured that subjects with minimal disease were not unnecessarily subjected to a risk of complications such as ptosis or facial weakness.

7.8.1.3. Study treatments

Subjects received either NT 201, reconstituted in 0.9% sodium chloride (NaCl), or matching placebo. In the main study period discussed here, subjects received a single treatment of up to 50 U per eye, in line with the standard recommendations for Botox treatment of blepharospasm. The dose was individualised per patient and was to be similar to the previous two injections prior to trial entry.

In the open label extension, subjects received up to 5 treatments, with dose adjustment as needed, depending on the effect achieved with the previous injection, in line with standard clinical practice.

7.8.1.4. Efficacy variables and outcomes

The main measurement tool for assessing blepharospasm was the Jankovic Rating Scale (JRS), which was used as the main outcome measure and also to ensure eligibility at baseline. The JRS ranges from 0-8 points. It consists of two items (subscores) "Severity" and "Frequency", with five rating categories each (0-4 points).

The JRS has been previously described and validated in the published literature and shows responsiveness to treatment.¹⁸ Its creator suggests that a change in the severity subscore of greater than or equal to 1 point is clinically meaningful.¹⁹

The primary efficacy variable was defined as the change from baseline in the JRS Severity subscore (JRSSS, assessed by a blinded Independent Rater) at Visit 4 (Week 6 plus or minus 3 days) after injection. This ranges from possible values of 0 to 4.

Secondary efficacy variables were listed as follows:

- Change from baseline in the JRSSS (assessed by Subject Diary) at Visit 4 after injection, based on the median score of the previous 7 days;
- Change from baseline in Blepharospasm Disability Index (BSDI) at Visit 4 after injection;
- Patient Evaluation of Global Response (PEGR) at Final Visit (Visit 5) of the Main Period.

Tertiary efficacy variables included:

- Change from baseline in the JRS Severity and Frequency subscores and in the JRS sum score (assessed by a blinded Independent Rater) at all other post-baseline visits;
- Change from baseline in the JRS Severity and Frequency subscores and in the JRS sum score (assessed by Subject Diary);
- Area under the Curve (AUC), time to maximum effect (T_{max}), and maximum effect (E_{max}) for change from baseline in the JRS Severity and Frequency subscores and in the JRS sum score (assessed by Subject Diary);
- Change from baseline in BSDI at all post-baseline visits except Visit 4;
- Global Assessment of Efficacy by Investigator (4-point Likert scale);
- Duration of treatment effect;
- Time to onset of treatment effect after injection;
- Time to waning of treatment effect after injection.

The Blepharospasm Disability Index (BSDI) is a scale specifically developed for assessment of the severity of BEB, and in particular it aims to measure how much BEB affects daily activities. It includes 6 items, each of which is to be assessed with a 5-point categorical response: “no impairment” (0 points), “slight impairment” (1 point), “moderate impairment” (2 points), “severe impairment” (3 points) and “no longer possible due to my illness” (4 points). The 6 items are “Driving a vehicle”, “Reading”, “Watching TV”, “Shopping”, “Walking” and “Doing everyday activities”. The BSDI allows subjects to answer “not applicable” for all items except “Doing everyday activities”.²⁰ The BSDI mean score is calculated by adding all applicable and answered items, and dividing by the number of items answered. A factorial analysis of the BSDI using principal components analysis has shown that one factor (principal component) is dominant and the mean score building is justified.²¹ It has also been shown to have acceptable retest-reliability.¹⁹ Overall, it appears to have appropriate properties for assessing the functional impacts of blepharospasm, and it is therefore suitable as a secondary efficacy

¹⁸ Bihari K, 2005; Iwashige H et al., 1995; Jankovic J, Orman J. 1987. Botulinum A toxin for cranial-cervical dystonia: A double-blind, placebo-controlled study. *Neurology* 37(4):616-623; Thussu A, Barman CR, Prabhakar S. 1999. Botulinum toxin treatment of hemifacial spasm and blepharospasm: Objective response evaluation. *Neurol India* 47(3):206-209; Roggenkämper P, Jost WH, Bihari K, et al. 2006. Efficacy and safety of a new Botulinum Toxin Type A free of complexing proteins in the treatment of blepharospasm. *J Neural Transm* 113(3):303-312.

¹⁹ Jankovic J, Kenney C, Grafe S, et al. 2009b. Relationship Between Various Clinical Outcome Assessments in Patients with Blepharospasm. *Mov Disord* 24(3):407-413.

²⁰ Goertelmeyer R et al, 2002; Grafe S et al, 2006.

²¹ Goertelmeyer R et al, 2002.

endpoint. Because of its subjective nature, however, it could be susceptible to significant placebo effects, and it is therefore appropriate that it was considered of secondary importance compared to the JRS.

The Patient Evaluation of Global Response (PEGR) is a subjective 9-point response scale that includes the following categories: very marked worsening (- 4), marked worsening (- 3), moderate worsening (- 2), slight worsening (- 1), unchanged (0), marked improvement (+ 1), moderate improvement (+ 2), slight improvement (+ 3), complete abolishment of signs and symptoms (+ 4). The scale was adapted from Jörg Wissel et al.²² No information was presented about the validity of this scale, but as it was a minor endpoint with a fairly transparent interpretation, this is acceptable.

7.8.1.5. Randomisation and blinding methods

Subjects were randomised in a 2:1 ratio to active treatment or placebo by a computer program (RANCODE, version 3.6) with stratification by centre. Blinding was achieved by having identical vials for active and placebo treatments, and maintained by keeping randomisation codes sealed and secured.

7.8.1.6. Analysis populations

The sponsor defined three study populations, the Evaluable for Safety population (EFS), the Intent-to-Treat population (ITT), and the Treated per-Protocol population (TPP), as in the previously described pivotal studies.

The ITT population was the primary population for confirmatory statistical analyses, including the primary efficacy endpoint.

7.8.1.7. Statistical methods

Analysis of the primary efficacy endpoint (change in JRS Severity subscore from baseline to Visit 4, Week 6) was based on the comparison of least square (LS) means from an ANCOVA model between the two treatment groups in the ITT population. The dependent variable was the change in the JRS Severity subscore assessed by the blinded Independent Rater, and the independent variables were treatment, baseline JRS Severity subscore, gender, age, dose group, and pooled centre. When data were missing, the last available value for the change was used (last observation carried forward (LOCF)). To check for the robustness of the results, the analysis was repeated with observed cases instead of LOCF imputation. A standard alpha value of 0.05 was used.

As in previous pivotal studies, full, final and simple ANCOVA models were developed. In this study, however, the full model was considered the primary model for confirmatory testing (unlike some other pivotal studies where the final model was considered primary).

Secondary endpoints were generally analysed with an ANCOVA performed on the ITT population. Analyses of E_{max} and T_{max} for displayed with Kaplan-Meier curves, and the two groups were compared with log-rank tests for equality over strata. For the global assessment of efficacy by the investigator, a descriptive Mann-Whitney test was used. An exploratory comparison between JRS subscores and other efficacy variables was performed with Pearson's correlation coefficient.

Overall, the statistical methods were appropriate and an ANCOVA was a suitable method for assessing the primary endpoint.

²² Wissel J, Müller J, Dressnandt J, et al. 2000. Management of spasticity associated pain with bo-tulinum toxin A. J Pain Symptom Manage 20(1):44-49.

7.8.1.8. Sample size

Although the sponsor intended to use an ANCOVA, they based power calculations on a two-group t-test with a two-sided significance level of $\alpha = 0.05$. (This simplifying approach is reasonable for a placebo-controlled study – if the study turned out to be under-powered, and significance was not achieved, no strong conclusions would be drawn; it would be inappropriate for an inferiority study). They determined that such a t-test would have 90% power to detect a difference greater than or equal to 0.8, assuming a common standard deviation of 1.0, for a sample size of 52 in the NT 201 group and 26 in the placebo group (with a randomisation ratio of 2:1). Allowing for a 30% drop-out rate, 75 subjects in the NT 201 group and 37 in the placebo group were to be randomised. Ultimately, these targets were not quite met (NT 201 75 randomised, placebo 34 randomised) but dropout was < 30% and the evaluable subjects exceeded the target. More importantly, significance was demonstrated for the primary endpoint, indicating that the study was adequately powered.

7.8.1.9. Participant flow

In the NT 201 group, 4 subjects were withdrawn because they met withdrawal criteria (planned total left knee replacement, interval between last and second-last Botulinum toxin injections prior to study start not > 10 weeks, subject not meeting JRS scale requirements, subject received open-label study medication) and one subject withdrew consent. In the placebo group, one subject was lost to follow up and one withdrew for unspecified reasons. Overall, there was no indication that the study was likely to have been compromised by withdrawal bias.

7.8.1.10. Major protocol violations/deviations

Protocol deviations are summarised in the table below. The overall frequency of violations was reasonably typical for a study of this nature. Withdrawals were slightly more common in the placebo group, but no consistent pattern was observed.

Table 42. Major protocol violations/deviations

Major Protocol Deviation [a]	NT 201 group	Placebo group	Total
	N = 75 n (%)	N = 34 n (%)	N = 109 n (%)
Total subjects with at least one deviation	4 (5.3)	3 (8.8)	7 (6.4)
Missing assessment	4 (5.3)	1 (2.9)	5 (4.6)
Administration of study medication	3 (4.0)	1 (2.9)	4 (3.7)
Missing Visit	3 (4.0)	1 (2.9)	4 (3.7)
Early discontinuation	3 (4.0)	2 (5.9)	5 (4.6)
Inclusion criterion at baseline not met: no need for injection of Botulinum toxin	1 (1.3)	2 (5.9)	3 (2.8)
Inclusion criterion at screening not met: no clinical diagnosis of BEB	1 (1.3)	0	1 (0.9)
Randomized subjects who were not treated	1 (1.3)	0	1 (0.9)
Wrong injection procedure	1 (1.3)	0	1 (0.9)

7.8.1.11. Baseline data

There were no important differences between the treatment groups at baseline. Disease severity, summarised in the subsequent table, was also similar at baseline, with JRS approximately 3. Some minor differences in BSDI were observed at baseline (mean 1.60 in the NT 201 group, 1.37 in the placebo group), but these are unlikely to have had a major impact on the study results. The ANCOVA model included baseline disease severity as an independent variable, which would be expected to compensate for minor baseline differences. The characteristics of the subjects' BEB history and the details of their previous Botox treatment prior to entry in the study were reasonably well-matched at baseline.

Overall, for a study of this size, matching was adequate, and the population studied appears typical of the intended target population.

Subjects had a range of concomitant diseases at baseline, with no important differences between the two groups.

7.8.1.12. Results for the primary efficacy outcome

In the NT 201 group, mean JRS Severity subscores decreased from 3.12 points at baseline to 2.29 points at Visit 4; in the placebo group Severity subscores increased from 2.94 points to 3.15 points. Subjects receiving active treatment showed a mean improvement of 0.83 points, compared to a mean deterioration of 0.21 points in the placebo group. This was significant by ANCOVA ($p < 0.001$). Note that the 95% CI in the sponsor's figure is missing a minus sign, and should read (-1.4 to -0.5), excluding zero. The subsequent table shows the correct 95% CIs for the primary analysis and for a number of supportive analyses, including analyses based on observed cases or imputation using group means instead of LOCF. These analyses, as well as similar analyses in the TPP population (not shown), suggested that the findings were robust and did not depend closely on the method of analysis.

7.8.1.13. Results for other efficacy outcomes

The major secondary endpoints includes changes in the total JRS score (sum score) at Week 6, changes in BSDI at Week 6 and the patient's evaluation of global response (PEGR). The p-values, derived from ANCOVA, were all statistically significant. The magnitude of the treatment effect as reflected in the BSDI was 0.8 points, which represents a substantial proportion of the baseline values: in the NT 201 group, the score decreased from a baseline of 1.60 to 1.20 points and in the placebo group, increased slightly from 1.37 to 1.47 points. This is consistent with a clinically worthwhile response. The PEGR showed a treatment effect of nearly 2 points (mean 1.9) from a potential range of 8 points (from -4 to +4); this is also likely to be clinically meaningful, as assessed by the patients themselves. Analysis of the TPP results produced very similar conclusions (not shown).

Tertiary efficacy variables (AUC and E_{\max} for JRS subscores) were consistent with the primary and secondary efficacy results, suggesting a robust treatment effect,

Assessment of T_{\max} showed a difference in the two groups, but the time to maximum effect lacks a clear interpretation in the placebo group, since no beneficial effect was observed and JRS scores showed a mean worsening. In the NT 201 group, the median T_{\max} for JRS Severity was observed at 5.5 weeks.

The global assessment of efficacy by the investigator showed a clear difference between the treatment groups, with poor responses in most placebo recipients and a range of outcomes in the active group, including about one third of subjects with "very good" responses and another third with "good" responses, as shown below. The treatment difference was significant by the Mann-Whitney test ($p < 0.001$).

In reference to the subjective times to onset of effect and the duration of effect inferred from re-injection intervals, onset was seen at a median of 4 days (mean 6.2) and the duration of effect had a mean of 74 days (mean 73.9). Comparisons with placebo were generally unhelpful for temporal measures, as the placebo group did not experience a meaningful clinical effect.

7.8.1.14. Subgroup analyses

The sponsor performed a range of subgroup analyses for the primary efficacy variable. The study was not powered for comparative statistics in each subgroup, and the results were presented descriptively. (The p-values in the table refer to the influence of the factor within the model, not to a between-treatment comparison.) Overall, this subgroup analysis was reassuring, and suggests that NT 201 is effective through most of the target population. Reductions (improvements) were seen with active treatment, regardless of gender, age and baseline JRS, but increases (deteriorations) were seen for most placebo subgroups. A minor improvement was seen in placebo subjects with baseline JRS of 4, but it was much less than the mean improvement seen in NT 201 recipients; the treatment difference in this group was consistent with the overall treatment difference.

7.8.2. Pivotal Study 0003

7.8.2.1. Study design, objectives, locations and dates

This blinded, randomised, parallel group, active-controlled study was titled “Safety and efficacy of NT 201 (highly purified Botulinum Neurotoxin A) compared to Botox (purified Botulinum Neurotoxin A –complex) in blepharospasm.” It was conducted in 42 centres in Belgium, Czech Republic, France, Germany, Hungary, Israel, Poland, Russia, Slovakia, from March 2001 to January 2002.

It was designed as a non-inferiority study, seeking to show that NT 201 was not worse than Botox in the treatment of blepharospasm, using doses up to 35 U per eye (70 U total per patient), in subjects previously treated with Botox.

7.8.2.2. Inclusion and exclusion criteria

Patients were eligible if they were adults of either gender aged up to 75 years, had BEB, had received Botox on at least two previous occasions and had shown a stable therapeutic response for the last two doses. A stable therapeutic response meant that the investigator and patient agreed the response had been the same to the two previous treatments, the doses were the same, and the difference in dose interval between treatments was no more than 3 weeks.

Subjects also had to be on stable doses of other medications used for dystonia, and females had to be able to avoid pregnancy.

Exclusion criteria included other significant diseases or contraindications to botulinum toxin, as listed for the previous pivotal studies.

A formal requirement for a specific severity of blepharospasm was not included in the inclusion or exclusion criteria, but to be eligible for Botox treatment even prior to study recruitment, subjects required significant blepharospasm.

7.8.2.3. Study treatments

Subjects received a single dose of up to 35 U of Botox or NT 201 in each eye, matched to their previous stable dose of Botox. They were then followed for up to 16 weeks.

7.8.2.4. Efficacy variables and outcomes

The primary efficacy variable was change from baseline in the JRS sum score, as assessed during a Control Visit on Day 21 plus or minus 1 day. Note that this is a slightly different endpoint to the JRS-Severity subscore which was used as the primary endpoint in the pivotal placebo-controlled study described above. For details of the JRS. This is a previously validated scale, and appropriate for use in this study. Use of the sum score in this study instead of the Severity subscore as in the previously described (but later performed) Study 0433 represents an arbitrary choice, but either measure seems appropriate.

Secondary efficacy variables were:

- Change from baseline in the JRS sum score at the Final Visit (approximately Day 109-112).
- Change from baseline in the mean total score for the “Function Scale for Patients with Blepharospasm” at the Control and Final Visits.
- Mean score for Patient Evaluation of Global Response (PEGR) at the control and final visits
- Global assessment of efficacy by the investigator
- Duration of treatment effect.
- Time to onset of treatment effect.
- Time to waning of treatment effect.

The “Function Scale for Patients with Blepharospasm” was identical to the Blepharospasm Disability Index (BSDI). The other measures were all as described for Study 0433. These are appropriate measure for assessing blepharospasm and its functional impact.

7.8.2.5. Randomisation and blinding methods

Subjects were block randomised to NT 201 or Botox in a 1:1 ratio using a software program, RANCODE. Blinding was maintained by preparing the study treatments off-site and supplying each treatment in identical-appearing syringes. The treatments do not have characteristic side effects distinct from their mode of action, so accidental unblinding was relatively unlikely.

7.8.2.6. Analysis populations

The sponsor defined three analysis populations, the EFS, ITT and TPP populations, as defined in previous pivotal studies.

The primary endpoint was assessed with the TPP population. This is outside the usual recommendations of the TGA, but may be preferable for a non-inferiority study, as discussed previously in this report. Because subjects only received a single treatment, the potential for withdrawal bias was limited and the TPP results would be expected to reflect the comparative efficacy of each drug. Also, as shown later, the ITT and TPP results were concordant.

7.8.2.7. Statistical methods

For the primary endpoint, the sponsor performed an ANCOVA with the change in JRS sum score as the dependent variable and the baseline sum score, dose, gender, age, number of previous injections, country and treatment*country interaction as independent variables. The final model was based on all variables with a moderate influence ($p < 0.2$) on the primary efficacy variable.

The mean difference between treatments for the primary efficacy variable was expressed as the least squares mean of the change from baseline for NT 201, minus that for Botox. Non-inferiority was to be inferred if the 95% CI for the difference was < 0.8 points.

The primary analysis was performed without data imputation, but the sponsor performed additional sensitivity analyses with different methods of handling missing data: an observed case analysis, an assumption of zero change from baseline, and imputation of the group mean.

Secondary efficacy variables were analysed with ANCOVA, except for the investigator’s global assessment of treatment efficacy, which was assessed with the Wilcoxon Rank-Sum test.

7.8.2.8. Sample size

The original power calculations were based on the assumption that a clinically meaningful change in the total JRS scale is 1 point. Jankovic et al, 2009 (publishing after this study was performed) suggested that a change of 1 point in the Severity subscore was clinically significant, implying that an even greater change was needed in the total sum score to imply clinical significance. A lower threshold of 0.8 points was suggested by the CPMP in December 2000, and this required recalculation of the necessary sample size.

In the original calculation, the sponsor aimed to achieve 90% power to demonstrate a difference of greater than or equal to 1.0 points, using a one-sided alpha level of 0.025, and estimated that 170 patients needed to be treated per protocol (214 randomised). In the revised calculation, the sponsor only aimed to achieve 85% power to demonstrate a difference of greater than or equal to 0.8 points, using a one-sided alpha level of 0.025. This required a sample size of 228 per-protocol patients (approximately 286 randomised patients).

The actual method of calculating sample size was somewhat unclear. The sponsor reports that nQuery Advisor was used to derive the sample size, but whether this estimate was performed with the ANCOVA in mind, and what assumptions were made (concerning the model, for instance, or the standard deviation of the efficacy variable), was not discussed. Given that this was a non-inferiority study – that is, one intended not to achieve statistical significance, the

power calculations are important in evaluating the results. In the eventual analysis of the results, though, the sponsor provided a 95% CI for the treatment difference between the two active treatments, and this fell well within the prespecified bound of a 0.8 point clinically meaningful difference. This indicates that, regardless of how it was determined prospectively, the sample size was actually adequate to show non-inferiority.

7.8.2.9. Participant flow

Patient disposition is summarised in the figure below. The difference between the TPP population and the ITT population was 44 subjects (NT 201 19, Botox 25), which is approximately 15% of the ITT population. This is reasonable for a study of this nature.

7.8.2.10. Major protocol violations/deviations

A review of the distribution of the protocol violations shows no concerning patterns. The most common violation was a failure to satisfy all of the inclusion criteria, particularly criterion 4, which referred to the need for stable doses, volumes, treatment intervals and therapeutic responses to the previous two Botox treatments. On balance, it seems unlikely that the violations substantially compromised the study.

Table 43. Protocol violations by patients enrolled in each treatment group

Protocol Violation TREATMENT CATEGORY	major		minor		no		Total	
	N	%	N	%	N	%	N	%
NT 201								
Time Window Violation	2	1.35	30	20.27	116	78.38	148	100.00
Course of Trial	4	2.70	41	27.70	103	69.59	148	100.00
ECG Irregularities	0	0	6	4.05	142	95.95	148	100.00
Screening Failure	0	0	0	0	148	100.00	148	100.00
Jankovic Rating	7	4.73	6	4.05	135	91.22	148	100.00
Incl. Crit. 2	0	0	5	3.38	143	96.62	148	100.00
Incl. Crit. 4	11	7.43	9	6.08	128	86.49	148	100.00
Incl. Crit. 5	0	0	0	0	148	100.00	148	100.00
Incl. Crit. 6	0	0	4	2.70	144	97.30	148	100.00
Incl. Crit. 7	0	0	2	1.35	146	98.65	148	100.00
Excl. Crit. 2	2	1.35	0	0	146	98.65	148	100.00
Excl. Crit. 5	0	0	0	0	148	100.00	148	100.00
Excl. Crit. 6	0	0	0	0	148	100.00	148	100.00
Injected Dose	1	0.68	1	0.68	146	98.65	148	100.00
Drug Irregularities	0	0	17	11.49	131	88.51	148	100.00
Random Order	0	0	16	10.81	132	89.19	148	100.00
Non-authorized Medication	0	0	1	0.68	147	99.32	148	100.00
No Safety Examination	4	2.70	0	0	144	97.30	148	100.00
BTXCo								
Time Window Violation	1	0.65	30	19.35	124	80.00	155	100.00
Course of Trial	3	1.94	49	31.61	103	66.45	155	100.00
ECG Irregularities	0	0	9	5.81	146	94.19	155	100.00
Screening Failure	3	1.94	0	0	152	98.06	155	100.00
Jankovic Rating	5	3.23	10	6.45	140	90.32	155	100.00
Incl. Crit. 2	0	0	6	3.87	149	96.13	155	100.00
Incl. Crit. 4	16	10.32	17	10.97	122	78.71	155	100.00
Incl. Crit. 5	0	0	1	0.65	154	99.35	155	100.00
Incl. Crit. 6	3	1.94	9	5.81	143	92.26	155	100.00
Incl. Crit. 7	0	0	2	1.29	153	98.71	155	100.00
Excl. Crit. 2	0	0	0	0	155	100.00	155	100.00
Excl. Crit. 5	1	0.65	0	0	154	99.35	155	100.00
Excl. Crit. 6	0	0	1	0.65	154	99.35	155	100.00
Injected Dose	2	1.29	5	3.23	148	95.48	155	100.00
Drug Irregularities	3	1.94	19	12.26	133	85.81	155	100.00
Random Order	0	0	10	6.45	145	93.55	155	100.00
Non-authorized Medication	1	0.65	1	0.65	153	98.71	155	100.00
No Safety Examination	2	1.29	0	0	153	98.71	155	100.00

7.8.2.11. Baseline data

Baseline features of the two treatment populations, including demographic and disease characteristics were acceptably matched in terms of severity and duration of blepharospasm. The total dose received by patients, which reflects their previous Botox dose, was also similar in the two groups.

7.8.2.12. Results for the primary efficacy outcome

Mean changes in the JRS sum score mean changes in each group were similar (NT 201 -2.83 points; Botox -2.65 points); the median change favoured NT 201 (NT 201 -3.0 versus Botox -2.0) but a comparison of medians is not very helpful, given that the discrete scoring system has steps greater than the difference between the group means. To basic visual inspection, the two treatments appear to be equal in efficacy.

Formal statistical analysis confirms that the groups were not significantly different. There was a slight trend in favour of NT 201, but the LS mean difference was small (-0.23) and the 95% CI included zero. The upper limit of the 95% CI for the difference included the possibility that Botox might be superior, but only by a trivial 0.21 points. This easily satisfies the prospectively defined no-difference limit of 0.8 points. Both groups showed a significant change from baseline ($p < 0.0001$), but without a placebo control group this does not constitute solid evidence of efficacy; some of the observed improvement could be due to a placebo effect or regression to the mean.

Similar results were obtained with other ANCOVA models. Results in the ITT analysis, were consistent with the TPP analysis. There was a trend in favour of NT 201 and even the upper limit for the 95% CI of the treatment difference was only consistent with trivial superiority of Botox (by 0.17 points).

Overall, these results convincingly demonstrate non-inferiority of NT 201 in comparison to Botox.

7.8.2.13. Results for other efficacy outcomes

The JRS Sum Scores at the final visit showed no important differences between the treatment groups, but there was a weak trend in favour of NT 201. The BSDI showed almost identical mean changes in the two groups at the control visit, but a trend in favour of NT 201 at the final visit which approaches statistical significance ($p = 0.06$ by ANCOVA). The PEGR also showed a weak trend in favour of NT 201. On balance, these findings are consistent with the two treatments being equivalent, supporting the analysis of the primary endpoint.

For the global assessment by the investigator, a weak trend in favour of NT 201 is apparent, with a slightly higher number of "very good" responses. This was not significant by the Wilcoxon Rank Sum test ($p = 0.14$ in the TPP analysis, $p = 0.07$ in the ITT analysis).

The mean duration of treatment effect, as estimated by the interval to the next dose, was almost identical in the two treatment groups, and no significant differences were found in a Cox proportional hazards model. The mean time to onset of treatment effect was slightly shorter with NT 201, and the mean time to waning was fractionally longer, but the differences were not significant.

7.8.2.14. Subgroup analyses

The sponsor stated that no subgroup analyses were performed with efficacy variables. Given that the two treatments appeared identical for the primary efficacy variable, and the study was only powered to show non-inferiority for the full cohort analysis, this is reasonable. Adequate exploration of the efficacy of NT 201 in various subgroups was explored in other submitted studies.

7.9. Evaluator's conclusions on clinical efficacy

Efficacy of NT 201 has been established for all four of the indications sought: glabellar frown lines, cervical dystonia, blepharospasm and spasticity. Pivotal studies in each indication showed highly significant superiority of NT 201 over placebo. Two pivotal Botox-controlled studies were also performed, one in cervical dystonia (0013) and one in blepharospasm (0003), and these showed non-inferiority of NT 201 according to pre-specified equivalence criteria.

All of the pivotal studies achieved their primary aim based on prospectively identified statistical methods, apart from the two pivotal studies of glabellar frown lines, where a zero response rate in the placebo group forced the sponsor to change statistical techniques. Secondary and tertiary endpoints across all four indications were consistent with the primary endpoints. On balance, the pivotal studies were well-conducted and free of significant methodological flaws, but one of the studies designated as pivotal (0607) merely compared two dilutions of NT 201 in the treatment of spasticity, and lacked a non-NT 201 treatment arm. It therefore does not provide robust evidence of efficacy, though efficacy in this indication was shown in another pivotal study (0410).

A range of supportive studies were also submitted, which were broadly consistent with the pivotal studies.

Repeat dose studies showed no evidence that the efficacy of NT 201 wanes with repeat dosing, but repeat doses were unblinded and uncontrolled, making it impossible to draw firm conclusions.

8. Clinical safety

8.1. Studies providing evaluable safety data

All of the studies described in this report contributed safety data, including the 8 pivotal studies, their open-label extensions, and the minor supportive studies. The three pharmacodynamic studies in healthy volunteers also contributed data, but the doses used were low. The safety database also including Study 0617, in which NT 201 was used to treat 'Crow's Feet' (CF) wrinkles. The sponsor is not seeking approval for this indication at present, so the study was not submitted in detail.

The sponsor identified two major pools of safety data in the Integrated Summary of Safety: the single-dose placebo-controlled studies, which provide the best evidence of side effects attributable to treatment, and the single-dose, active-controlled studies, which allow comparison of the side effect profile of NT 201 in relation to competing products. The combination of these two pools was also analysed. Repeat-dose studies were also assessed, but this pool of data is less useful because it lacks a comparator.

A total of 2,068 patients were treated with NT 201 in single-dose studies, and 1,313 of these also entered repeated-dose studies.

Safety monitoring was typical of a large clinical program for a new agent, and was acceptable overall. In all studies, subjects and their investigators reported adverse events (AEs) at regular scheduled visits. In addition, any unscheduled clinic or hospital attendances were noted. Narrative summaries were provided for any major events.

For NT 201, which is administered locally, the safety implications depend strongly on the site of injection. In particular, given the known side effect profile of botulinum toxin, which includes unwanted weakness in injected muscles or neighbouring muscles, the sponsor specifically searched the AE database for events suggestive of toxin-induced weakness. Other complications directly related to the interruption of cholinergic transmission include dry eyes and dry mouth, which were also considered AEs of special interest.

Routine laboratory monitoring was performed in all studies. All major studies included an assessment of the immunogenicity of NT 201.

8.2. Patient exposure

Patient exposure to NT 201 in the pooled single-dose studies is summarised below. A total of 2,068 patients were treated. More than half of these (1,067) received NT 201 for GFL, and therefore received a low dose (range 10-30 U). Higher doses were tested in the CD population (n = 431, median dose 120 U) and in the treatment of spasticity (n = 265, median dose 300 U). Medium doses were administered in subjects with blepharospasm (n = 222, median dose 50 U). The number of subjects with Crow's Feet (CF) was low (n = 83); this population received a dose (24 U) similar to that used in GFL.

Cumulative exposure was more substantial in the repeated-dose studies, summarised below. Over a thousand patients (1,313) were involved in repeated-dose studies, and all of the major sites and indications were assessed.

8.3. Adverse events

8.3.1. All adverse events (irrespective of relationship to study treatment)

8.3.1.1. Single-dose placebo-controlled studies

Treatment-emergent adverse events (TEAEs) in the single-dose placebo-controlled studies are summarised in the table below. Compared to placebo, there was an overall excess of TEAEs in NT 201 recipients of 9.4% (46.4% - 37.0%). For TEAEs of severe intensity, the attributable excess was 3% (5.9% - 2.9%).

Table 44. Summary of TEAEs in the pooled placebo-controlled single-dose studies by treatment.

	NT 201 N=1067	Placebo N=527
Subjects with TEAEs	495 (46.4%)	195 (37.0%)
Subjects with severe TEAEs	63 (5.9%)	15 (2.9%)
Subjects with related TEAEs	180 (16.9%)	38 (7.2%)
Subjects with SAEs	18 (1.7%)	4 (0.8%)
Subjects with related SAEs	0	0
Subjects who died ¹	0	1 (0.2%) ¹
Subjects with TEAEs leading to dropout	5 (0.5%)	1 (0.2%)
Subjects with AESIs	78 (7.3%)	12 (2.3%)

¹ Six deaths have occurred in the completed studies of NT 201. See Section 5.3.1.2.1 below for more details.

Data are presented as number of subjects (percent of subjects).

AESI: adverse event of special interest; BLEPH: blepharospasm; CD: cervical dystonia; CF: crow's feet; GFL: glabellar frown lines; SAE: serious adverse event; SP: spasticity; TEAE: treatment emergent adverse event.

The pooled placebo-controlled studies were: 0520 Main Period, 0527 (GFL), 0724 (GFL), 0741 (GFL) 0617 (CF), 0408/1 Main Period (CD), 0433/1 Main Period (BLEPH), 0410/1 Main Period (SP).

Some of the most common TEAEs occurred with similar or greater frequency in the pooled placebo group, and it appears unlikely that NT 201 played a causal role: nasopharyngitis and other infections, in particular, were common in both groups.

The TEAEs that showed an excess in the NT 201 group included: headache, musculoskeletal disorders (including neck pain, back pain and musculoskeletal pain, and weakness), gastrointestinal disorders (including dysphagia), oropharyngeal pain, eyelid ptosis and dry eye. For most of these, a plausible causal relation between active treatment and the TEAE can be envisaged, particularly TEAEs that involve weakness (ptosis, dysphagia) or impairment of secretions (dry eye, dry mouth). Some of the other categories showed only small differences between the active and placebo groups, so a causal relation seems less likely. Procedural complications and injection site pain were slightly more common with active treatment.

TEAEs according to the dose of NT 201 administered, unfortunately, does not produce a clear idea of the dose-dependence of adverse events, because the dose was different for each indication, and the dose-group therefore acts as a surrogate marker for injection site. For instance, injecting a high dose into the upper limb is unlikely to produce eyelid ptosis, whereas injecting near the eye is much more likely to produce ptosis, even at low doses; this creates the

misleading appearance of a lower incidence of ptosis at higher doses. When controlled for site, it would be expected that all weakness-related AEs showed an increased incidence at higher doses. Instead, site-specific safety needs to be considered.

8.3.1.2. Other single-dose studies

In the overall pooled single-dose studies, including the active-controlled studies as well as the placebo-controlled studies, the incidence of TEAEs appeared similar with NT 201 and placebo, and the incidence of TEAEs with both NT 201 and placebo were higher than with Botox.

This is likely to reflect that the mix of indications was not the same in the different pools. All four of the proposed indications were assessed with a placebo-controlled design, including several minor studies and five of the eight pivotal studies: GFL (two pivotal studies), CD (one pivotal study), blepharospasm (one pivotal study) and spasticity (one pivotal study). The Botox-controlled studies only included two pivotal studies: CD (one pivotal study) and blepharospasm (one pivotal study). Two additional non-pivotal studies had active controls (Study 3002, GFL; Study 9801, CD). One additional pivotal study in patients with spasticity was performed with two different dilutions of NT 201, and thus did not contribute placebo-controlled or Botox-controlled safety data.

Because of the differences in the ages, comorbidities, injection sites and doses associated with each indication, comparing drugs across the entire pool of all these studies is not a valid exercise. A more direct comparison between NT 201 and other botulinum toxin preparations can be obtained from the four individual active-controlled studies:

- Supportive Study 3002 (GFL, Vistabel);
- Pivotal Study 0013/1 and dose-ranging Study 9801 (CD, Botox);
- Pivotal Study 0003/1 (BLEPH, Botox).

The AEs obtained in individual active-controlled studies are listed in the following pages. Overall, the incidence of AEs was similar with NT 201 and other botulinum toxin A preparations, as would be expected given that Botox can be considered a pro-drug for NT 201, and dissociates soon after injection to release purified toxin identical to NT 201.

8.3.1.2.1. AEs in Study 3002 (NT 201 24 U versus Vistabel 24 U for GFL)

The incidence of AEs was moderate, but the most common AE category was infections, which appear unlikely to be causally related to treatment. The overall incidence of AEs was slightly higher with Vistabel than with NT 201, but the pattern of AEs was similar with the two agents.

8.3.1.2.2. AEs in Pivotal Study 0013. 70-300 U of NT 201 or Botox, for Cervical Dystonia

In this study, AEs were slightly more common with NT 201 than Botox, but the general pattern of AEs was similar with the two agents, and raised no new safety concerns.

8.3.1.2.3. AEs in Study 9801. NT 201, total 30-90 U versus Botox, total 90 U for Cervical Dystonia

This study was underpowered for a safety analysis, with just 12 Botox recipients, so the incidence of different AEs was highly variable across doses and treatments. Overall, the study did not raise any significant safety concerns.

8.3.1.2.4. AEs in pivotal Study 003, NT 201 versus Botox for Blepharospasm

In this study, the incidence of AEs was similar for both active treatments, but marginally higher for Botox. Serious adverse events were rare, but they were slightly higher with Botox. Individual AEs occurred with a similar incidence with the two drugs.

Overall, there are no consistent differences between the active treatments.

8.3.2. Repeated-dose studies

In the repeated-dose studies, the incidence of TEAEs was higher than in single-dose studies, which is likely to reflect, in part, the longer period of follow-up and the occurrence of ailments that are common in the general population. Without a control group, it is difficult to draw any conclusions from this data.

When the TEAEs from repeated-dose studies are categorised according to which injection session they occurred in, no concerning pattern emerges. There is an overall trend to less TEAEs for later injections, compared to earlier injections, but this could reflect the fact that subjects and investigators failed to report events similar to those already reported for earlier injections, or that they lost enthusiasm for reporting events as the study proceeded. Without a control group, the data is of limited value.

The individual TEAEs reported in repeated-dose studies resemble those already considered in single-dose studies. Infections were common, especially nasopharyngitis, but this is likely to reflect the fact that such infections are common in the general community. The other TEAEs listed below were more common with NT 201 than placebo in the single-dose studies, and their continued occurrence in the repeated-dose studies is therefore likely to reflect, in part, a causal contribution from NT 201. This particularly applies to events suggestive of weakness (muscular weakness, dysphagia, ptosis) or impaired glandular secretions (dry mouth, dry eye).

TEAEs in repeated-dose studies by dose group, (as discussed earlier) present difficulties in assessing the data because dose group in the pooled population is a surrogate marker for site and indication.

8.4. Treatment-related adverse events (adverse drug reactions)

8.4.1. Single-dose placebo-controlled studies

Adverse events in which a causal role of NT 201 was suspected by the reporting investigator were classified as 'adverse reactions'. Inferring a causal role in individual cases is inherently unreliable, because it may merely reflect the expectations of the investigators – unexpected or novel adverse events might not be recognised as related to treatment. On the other hand, the investigator has a chance of viewing the temporal relationship between the AE and treatment, a relationship often not captured in tables of TEAEs. Also, causal attribution for studies of NT 201 is likely to have been improved by the investigators' previous experience with other botulinum toxin preparations. Overall, a list of 'adverse reactions', albeit imperfect, gives some indication of the side effect profile of the drug. The data suggests that headache, weakness, musculoskeletal pain, dysphagia, dry mouth, dry eye and ptosis were all attributed to active treatment more often than to placebo. The absolute incidence of most of these individual reactions was low, however, with only headache and dysphagia exceeding 2%.

8.4.2. Other studies

Adverse reactions in the repeated-dose studies were quite common, and this varied strongly with different indications. There was no evidence of an increasing incidence of adverse reactions with repeated treatment; instead there was a trend to report less reactions as the studies progressed.

8.5. Serious adverse events

8.5.1. Single-dose placebo-controlled studies

In single-dose placebo-controlled studies, serious adverse events (SAEs) occurred in 18 of 1067 NT 201 recipients (1.7%), and 4 of 527 placebo recipients (0.8%). Although based on low numbers, these raw incidences suggest that NT 201 may have played a causal role in some SAEs;

another 4 SAEs (a doubling of the observed SAEs) would be needed for the placebo incidence to match the NT 201 incidence.

A review of individual SAEs, however, shows no case in which a likely causal relation can be envisaged between the local toxin treatment and the distant site of the problem. For instance, many SAEs involved problems in the pelvis after injection of a low dose in the forehead, or problems in the lungs after injection in the limbs. In no case did the investigator indicate that the SAE was related to treatment, and this suggests that it is unlikely that the investigators missed a hidden causal relationship.

There was one case of pyrexia which occurred in a patient with GFL; this is unlikely to have represented an immune response to the drug because the fever developed almost ten weeks after injection, and responded to antibiotics. Of the cases that were suggestive of muscle weakness (such as pelvic floor muscle weakness in a GFL patient), the timing of the weakness and the dose administered makes it very unlikely NT 201 contributed.

8.5.2. Other studies

A review of SAEs in other studies, including the active-controlled and repeated-dose studies, did not raise new concerns. The incidence of SAEs was slightly higher in Botox recipients than in NT 201 recipients, but a direct comparison is not valid because the mix of indications contributing to the pools was unequal. SAEs in individual active-controlled studies did not suggest that there were any important differences between NT 201 and Botox.

Repeated-dose studies showed a higher incidence of SAEs, consistent with the longer period of observation, but the incidence did not increase with successive injections.

8.6. Deaths

Deaths in the NT 201 study program were rare. There were six known deaths at the time of submission. One of these was in a placebo recipient. For the other deaths, the nature of the events was consistent with the comorbidities of the target population. Four deaths occurred in subjects being treated for post-stroke spasticity, and consisted of vascular complications that are common in this population. One subject treated for cervical dystonia developed colon cancer. Overall, a causal role of NT 201 seems unlikely in all cases.

8.7. Discontinuation due to adverse events

8.7.1. Single-dose studies

Discontinuations due to AEs were relatively rare. In the single-dose placebo-controlled studies, TEAEs leading to dropout occurred in 5 NT 201 recipients (0.5%) and 1 placebo recipient (0.2%). In the no placebo controlled studies, some additional dropouts occurred, including two in Botox recipients.

The individual events are summarised below.

For the three subjects with cervical dystonia who dropped out of Study 0408 after receiving NT 201, a possible causal relationship was indicated by the investigator. For the two subjects with muscular weakness, a causal role appears plausible; for the case of nausea and dizziness, it is unclear if NT 201 contributed. TEAEs in other studies appear relatively unlikely to have been caused by NT 201.

8.7.2. Other studies

In the repeated-dose studies, 14 of 1313 subjects (1.1%) reported TEAEs leading to discontinuation. Individual AEs classified by preferred term were single occurrences with the exception of dysphagia, dry mouth, and urinary tract infection, each of which occurred in 2

subjects. Dysphagia and dry mouth are known complications of botulinum treatment in the neck and are plausibly related to treatment.

8.8. Adverse events of special interest

Several AEs are known to occur after treatment with botulinum toxin, including several related to unwanted weakness as a direct pharmacological response to the drug. Autonomic complications may also occur; in particular, glandular secretions may be inhibited, leading to dry eyes or dry mouth.

The sponsor searched for AEs falling into these categories, and performed a statistical assessment of the incidence in the active and placebo groups, grouped by indication. The results were presented for single-dose placebo-controlled studies, and for all studies, as displayed in the following sections.

8.8.1. Single-dose placebo-controlled studies

AEs of special interest (AESIs) in single-dose placebo-controlled studies are listed below, for each indication. There was an excess of AESIs in the NT 201 groups, and this was most marked for subjects with cervical dystonia, who had an attributable AESI incidence of 16.6% (95% CI 8.4% to 24.9%). The attributable incidence of AESIs was also elevated in subjects with blepharospasm, but the 95% CI was not quite statistically significant (difference, 17.3%, 95% CI -0.3% to 34.8%). Several individual AESIs showed a significant difference in incidence for NT 201 recipients and placebo recipients, at least for some indications.

The most common AESIs were dysphagia, eyelid ptosis, muscular weakness and dry mouth. A statistically significant difference ($p < 0.05$) between NT 201 and placebo was observed for the overall incidence of AESIs, gastrointestinal AESIs, eye AESIs, , musculoskeletal/connective-tissue AESIs, and nervous system AESIs, as well as for the specific events of dysphagia, dry mouth, eyelid ptosis, muscular weakness, and facial paresis.

All of these AEs are consistent with the known adverse event profile of botulinum toxin, and are intrinsically tied in with the toxin's intended mode of action. NT 201 does not appear to pose any new risks in relation to these problems.

8.8.2. Other studies

A similar search for AESIs was performed in the complete safety database, including non-placebo-controlled studies. The results are shown below. The same problems were identified, and no significant new safety concerns were raised. As in the single-dose placebo-controlled studies, the conditions associated with the highest incidence of AESIs were cervical dystonia and blepharospasm. In both conditions, this reflects the difficulty in finding a balance between weakening muscles just enough to reduce symptoms, but not so much that they become functionally compromised, as well as the difficulty in limiting the effects of the toxin purely to the targeted muscles and not the neighbouring structures.

The evidence from the repeated dose studies suggests that reports of AESIs do not increase with repeated injection, but instead decrease. Without a control group, no firm conclusions can be drawn from this observation, but some improvement could be due to optimising the dose in response to AESIs observed in the previous treatment cycle. It is also possible that patients and investigators failed to report problems that had already been documented in earlier cycles.

8.9. Immunogenicity

The immunogenicity of NT 201 can be considered in two contexts: the development of neutralising antibodies, which potentially compromises the efficacy of the toxin, and the occurrence of clinically significant allergic reactions.

In many of the submitted studies, subjects had previously received other commercial preparations of botulinum toxin and already had anti-botulinum toxin antibodies. For subjects in single-dose studies who were antibody-negative at screening, the incidence of positive antibody results on a fluorescence immunoassay (FIA) the incidence of neutralising antibodies in a mouse hemi-diaphragm assay (HDA) is shown in the last column. Neutralising antibodies were slightly more common with Botox (0.7%) than with NT 201 (0.4%), but the analysis was not adequately powered for a statistical comparison. Neutralising antibodies did not develop in the placebo group.

Many of the antibodies seen in NT 201 recipients were transient.

This data is broadly consistent with pre-clinical studies, which suggested that the immunogenicity of NT 201 is lower than with other forms of botulinum toxin.

Occasional adverse events were reported in antibody-positive subjects, but none of the treatment-related AEs in this subgroup were suggestive of immune-mediated events, such as hypersensitivity reactions or anaphylaxis.

When all allergic AEs are included, including those thought unrelated to treatment, treatment with NT 201 appeared to be associated with a low incidence of allergic symptoms, and in most cases a causal relation between treatment and the symptoms was uncertain.

The following allergic or immune-mediated symptoms were listed in the sponsor's Risk Management Plan.

The incidence in repeated-dose studies was similar:

'Repeated dose studies: Rash 0.81% (3 of 369 subjects); all mild; all not related, all non-serious; Dermatitis contact 0.27% (1 of 369 subjects); moderate; not related; non-serious; Urticaria 0.54% (2 of 369 subjects); 0.27% mild, 0.27% moderate; both not related, both not serious; Papular rash 0.27% (1 of 369 subjects); moderate; not related; non-serious; Conjunctivitis allergic 0.81% (3 of 369 subjects); 0.54% mild; 0.27% moderate; all not related; all non-serious; Rhinitis allergic 1.08% (4 of 369 subjects); 0.27% mild; 0.81% moderate; all not related; all non-serious; Hypersensitivity 0.54% (2 of 369 subjects); 0.27% mild; 0.27% moderate; both not related; both non-serious; Dermatitis allergic 0.27% (1 of 369 subjects).'

The sponsor also notes that "local allergic reactions and flu-like symptoms" have emerged as a safety signal during post-marketing surveillance, although the nature and strength of this safety signal was not discussed in the submission. This should be clarified by the sponsor.

8.10. Laboratory tests

In the sponsor's Integrated Summary of Safety, mean changes in laboratory results were not summarised in a convenient tabular format. For a start, the relevant tables were contained in appendices that were missing from the original digital submission. Although these missing appendices were supplied upon request, the mean shift for each laboratory parameter was shown in its own table, so that the data ran for more than a hundred pages. A review of these tables did not reveal important differences between NT 201, placebo and Botox (data not shown), but the lack of appropriate summary tables makes it difficult to draw firm conclusions about the safety of NT 201.

The incidence of clinically significant shifts in laboratory results was presented in a more appropriate manner. , clinically significant changes in the overall pooled single-dose studies were seen with a comparable incidence in the pooled NT 201 (47 of 2068, 2.27%) and pooled placebo (17 of 527, 3.23%) groups. The incidence in the pooled Botox group was lower (3 of 493, 0.61%), but this was based on a low overall number of Botox recipients. A review of the individual categories of clinically significant laboratory abnormalities showed that the most

common abnormality in the pooled NT 201 group was high glucose, but the incidence of this abnormality was actually slightly higher with placebo than with NT 201.

In the repeated dose studies, no significant safety concerns emerged in relation to biochemistry monitoring. The sponsor summarised the abnormalities as follows:

'In the repeated-dose studies, 99 of 1313 subjects (7.5%) had in total 159 treatment-emergent, clinically significant biochemistry abnormalities. These events were elevated total cholesterol (34 subjects, 2.6%), elevated LDL cholesterol (23 subjects, 1.8%), elevated glucose (22 subjects, 1.7%), elevated GGT (21 subjects, 1.6%), elevated ALT (10 subjects, 0.8%), low HDL cholesterol, elevated AST (9 subjects, 0.7%), elevated serum urea nitrogen (7 subjects, 0.5%), low glucose, elevated alkaline phosphatase (5 subjects, 0.4%), elevated potassium (3 subjects, 0.2%), elevated bilirubin, elevated calcium, low LDL cholesterol (2 subjects, 0.2%), elevated HDL cholesterol, elevated creatinine, low potassium, low and elevated sodium (1 subject, 0.1%).'

These incidences are broadly within expectations for a middle-aged population monitored for several months, but it is difficult to draw conclusions in the absence of a control group. Given that Botox has been widely used for many years without significant biochemistry abnormalities emerging as a safety concern, and that NT 201 contains the same active component as Botox, it seems unlikely that NT 201 poses substantial risks of causing clinically significant biochemical disturbances.

8.11. Liver function

The previous experience with Botox suggests that NT 201 would be expected to have minimal risk of causing hepatic abnormalities, and the submitted data for NT 201 support this.

The incidence of abnormal parameters on liver function tests was low (SGOT/ASAT was high in 8 NT 201 recipients, GOPT/ALAT high in 7 NT 201 recipients, and bilirubin high in 3 NT 201 recipients). Abnormal liver function results were not seen in the placebo group, but the placebo group was relatively small, and the data set may have lacked sufficient power to characterise the low background placebo incidence of abnormal liver function tests.

Hepatic AEs did not feature in a list of treatment-emergent AEs reported in the pooled placebo-controlled studies.

8.12. Kidney function

In the pooled single-dose studies, the incidence of clinically significant abnormalities of serum creatinine was low, with just one patient affected, an NT 201 recipient (1/2068, 0.05%). No placebo or Botox recipients had clinically significant elevations of creatinine. Clinically significant elevations of urea were also rare, with four NT 201 recipients affected (4/2068, 0.19%) and no placebo or Botox recipients. Among the treatment-emergent AEs reported in the pooled placebo-controlled studies, no events related to renal impairment were listed.

8.13. Other clinical chemistry

As discussed above, the incidence of biochemistry abnormalities in the submitted studies was low, and no worse in the pooled NT 201 group than the placebo group.

8.14. Haematology

The results of haematological monitoring were not submitted in a convenient format, and the mean changes for individual haematological parameter were displayed in separate tables that

were not included in the original digital submission. A review of these mean changes did not suggest that there were important differences between NT 201 and placebo (data not shown).

Clinically significant haematological abnormalities were rare in all treatment groups, and there were no important differences in the incidence of clinically significant abnormalities across groups. In the overall pooled single-dose studies, including active-controlled studies, 11 of 2068 NT 201-treated subjects (0.5%) had treatment-emergent, clinically significant haematology abnormalities, compared to 2 of 493 subjects in the Botox group (0.4%) and 5 of 527 subjects (1.0%) in the placebo group.

In repeated dose studies, no important changes in mean haematological parameters occurred, and the emergence of clinically significant abnormalities occurred at a low and acceptable rate, summarised by the sponsor as follows:

'In the repeated dose studies, 34 of 1313 subjects (2.6%) had in total 67 treatment-emergent, clinically significant haematology abnormalities. The abnormalities were low neutrophils (10 subjects, 0.8%), low leucocytes (7 subjects, 0.5%), low haemoglobin (6 subjects, 0.5%), low haematocrit, low monocytes, low platelets, low reticulocytes (promille), low erythrocytes (5 subjects each, 0.4%), elevated platelets (4 subjects, 0.3%), low reticulocytes (109/L), low lymphocytes (3 subjects, 0.2%), elevated eosinophils, elevated leucocytes, elevated neutrophils (2 subjects, 0.2%), elevated basophils, elevated reticulocytes (both concentration and promille) (1 subject each, 0.1%).'

Overall, the use of NT 201 appears to have an acceptable haematological risk.

8.15. Electrocardiograph

8.15.1. Single-dose placebo-controlled studies

Not all single-dose studies performed a post-treatment ECG; some performed an ECG at screening only. For those studies in which a post-treatment ECG was performed, there was no evidence of any systematic change in any of the major ECG parameters.

On balance, given the mode of action of NT 201, and the extensive post-marketing experience with botulinum toxin A, it would not be expected that local injections at the sites tested would produce distant effects on the myocardium.

8.15.2. Other studies

The sponsor performed a detailed analysis of ECGs for two active-controlled studies: pivotal Study 0003, in blepharospasm, and pivotal Study 0013, in cervical dystonia. These represent the two major studies in which NT 201 was compared to Botox. As shown in the tables below, there were no important differences between the two treatments in mean ECG parameters at baseline, at the assessment ('Control') visit, and the final visit. Minor changes in the QT interval (increases > 30msec) were observed in a small proportion of patients, but these appeared no more likely at the Control visit, when drug effect would be expected to be maximal, than the final visit. On balance, this evidence does not suggest that NT 201, when used as indicated, poses new or significant cardiac risks.

8.16. Vital signs

The overall results of vital-sign monitoring were not presented in a convenient format: the Integrated Summary of Safety referred to summary tables that were contained in appendices omitted from the original digital submission. When obtained, the shift tables summarised each parameter (systolic blood pressure, diastolic blood pressure, heart rate, respiratory rate) and each treatment (NT 201, Botox and placebo) in separate tables, making it difficult to present the data in this evaluation report. Overall, there did not appear to be any important differences between NT 201 and the active or placebo control in the incidence of abnormal vital signs.

The tables below summarise the univariate statistics for mean changes in each parameter, by treatment group and visit. There do not appear to be any important differences between treatments.

8.17. Post-marketing experience

In addition to extensive post-marketing experience with closely related products, such as Botox, there is also some post-marketing data available for NT 201, which was approved in Europe in 2005.

According to the sponsor:

'Since the launch of the product in Germany on 01 July 2005, approximately 127,500 subjects worldwide have been treated with NT 201. Treatment with NT 201 shows good efficacy, and is safe and well-tolerated. For current safety data, please refer to the most recent Periodic Safety Update Report (PSUR). To date, the post-marketing safety analysis has not revealed any new safety signals except for local allergic reactions and flu-like symptoms.'

The nature of these safety signals was not discussed in the sponsor's Integrated Summary of Safety, and the sponsor should be asked to clarify this.

8.18. Safety issues with the potential for major regulatory impact

8.18.1. Liver toxicity

As discussed, the incidence of hepatic dysfunction was low in the submitted studies. Liver toxicity is not a feature of treatment with other botulinum toxins, and there does not appear to be a significant risk of hepatotoxicity with NT 201.

8.18.2. Haematological toxicity

The haematological data do not suggest that NT 201 has any significant effects on bone marrow function or cell counts.

8.18.3. Serious skin reactions

Serious skin reactions were not reported as TEAEs. According to the sponsor, local allergic reactions have emerged as a safety concern during post-marketing surveillance, but the extent of the problem is unclear from the sponsor's submission. This should be clarified.

8.18.4. Cardiovascular safety

There is no evidence of significant cardiovascular risk with NT 201 treatment. Vital signs and ECG monitoring did not show significant differences between NT 201 and placebo, and cardiovascular AEs occurred at the expected rate in NT 201 recipients.

8.18.5. Unwanted immunological events

Immunological events occurred at a low incidence in NT 201 recipients and placebo recipients, and were generally thought by investigators to be unrelated to treatment. According to the sponsor, however, flulike symptoms and local allergic reactions have emerged as a safety signal during post-marketing surveillance. The nature of this safety signal should be clarified.

8.19. Other safety issues

8.19.1. Safety in special populations

Botulinum toxin preparations should be used with extreme caution in subjects with significant neuromuscular weakness, because a worsening of pre-existing weakness would be expected in

any muscle region exposed to the toxin. In this respect, NT 201 does not pose new safety concerns compared to competing botulinum preparations.

The safety of NT 201 in different subgroups is summarised below. Overall, the safety of NT 201 appears acceptable in both younger and older patients, in both males and females, and in pre-treated and treatment-naïve subjects.

A more detailed breakdown of the incidence of AEs by age and gender is provided in the tables below. AEs were more common in subjects older than 65 years, as expected, but no significant new safety concerns are posed by treatment in this age group.

The safety of NT 201 in pregnant women is unclear. According to the sponsor's Risk Management Plan (RMP), 15 reports of drug exposure to NT 201 before or during pregnancy had been received, with no evidence of teratogenesis. Seven pregnancies ended with a healthy birth, one was electively terminated, two ended with spontaneous abortions and another two were missed abortions. One pregnancy report was due to a blighted ovum. The sponsor proposes a pregnancy rating of category B3, which seems appropriate.

8.19.2. Safety related to drug-drug interactions and other interactions

Because NT 201 and other botulinum preparations can cause weakness, they should be used with caution when combined with other agents that can cause weakness, in particular anti-spasmodic agents. Pharmacokinetic interactions with NT 201 are expected to be minimal, however, because the drug is not used systemically and appreciable serum concentrations are not achieved.

8.20. Evaluator's overall conclusions on clinical safety

Overall, the safety of NT 201 is acceptable. The main risks associated with treatment are those already faced by patients being treated with existing botulinum toxin preparations. The most significant toxicity issue is the development of excessive weakness in targeted muscles or unwanted weakness in neighbouring muscles. This can be manifested as eyelid ptosis, following injection near the eye, dysphagia or dysarthria, following injection in the neck, or limb weakness following injection in the limbs. Interference with autonomic function may also occur, leading to dry eyes or dry mouth, though the incidence of these problems was low in the submitted studies. These side effects are intrinsically related to the drugs mode of action, and can be minimised by careful dose titration and by restricting use of the drug to experienced operators.

Other safety concerns arise from the fact that NT 201 is a foreign protein, and has some immunogenicity, though this appears less than with other botulinum toxin preparations. The sponsor mentions that flu-like symptoms and local allergic reactions have emerged as safety signals during the post-marketing experience of this drug, and this should be clarified. Such symptoms were not prominent during the submitted studies, so it seems unlikely that this will turn out to be a major safety issue.

9. First round benefit-risk assessment

9.1. First round assessment of benefits

The benefits of NT 201 in the proposed usages are:

- Reduction in glabellar frown lines, with a response rate of at least 54% compared to 0% with placebo;
- Amelioration of cervical dystonia that is clearly superior to placebo and equivalent to that achieved with Botox;

- Reduction in blepharospasm that is clearly superior to placebo and equivalent to Botox;
- Reduction in post-stroke spasticity that is superior to placebo

9.2. First round assessment of risks

The risks of NT 201 in the proposed usage appear to be the same as with existing botulinum preparations, and the registration of NT 201 would not be expected to pose any significant new risks.

The main complications of NT 201 are those related to excess weakness in the injected muscles or neighbouring muscles, and dry eyes or dry mouth.

The sponsor reports that post-marketing surveillance has revealed some risk of allergic responses, but the details of this risk have not been well characterised. Unwanted immunological events in the submitted studies were rare.

9.3. First round assessment of benefit-risk balance

The benefit-risk balance of NT 201, given the proposed usage, is favourable.

10. First round recommendation regarding authorisation

NT 201 should be approved for the proposed indications.

11. Clinical questions

11.1. Pharmacokinetics

Not applicable.

11.2. Pharmacodynamics

Not applicable.

11.3. Efficacy

Not applicable.

11.4. Safety

The RMP mentions that post-marketing surveillance has identified “ “. The sponsor should be asked to clarify this.

12. References

Albright AL, Barron WB, Fasick MP, et al. 1993. Continuous intrathecal baclofen infusion for spasticity of cerebral origin. JAMA 270(20):2475-2477.

Allergan Ltd. BOTOX . Botulinum Toxin Type A - Summary of Product Characteristics. June 25 2007.

- Anderson RL, Patel BC, Holds JB, et al. 1998. Blepharospasm: Past, present, and future. *Ophthalmol Plast Reconstr Surg* 14(5):305-317.
- Antonucci F, Rossi C, Gianfranceschi L, et al. 2008. Long-distance retrograde effects of botulinum neurotoxin A. *J Neurosci* 28(14):3689-3696.
- Ascher B, Zakine B, Kestemont P, et al. 2004. A multicenter, randomized, double-blind, placebo-controlled study of efficacy and safety of 3 doses of botulinum toxin A in the treatment of glabellar lines. *J Am Acad Dermatol* 51(2):223-233.
- Ascher B, Zakine B, Kestemont P, et al. 2005. Botulinum Toxin A in the Treatment of glabellar lines: Scheduling the next injection. *Aesthetic Surgery Journal* 25(4):365-375.
- Bakheit AM, Thilmann AF, Ward AB, et al. 2000. A randomized, double-blind, placebo-controlled, dose-ranging study to compare the efficacy and safety of three doses of botulinum toxin type A (Dysport) with placebo in upper limb spasticity after stroke. *Stroke* 31(10):2402-2406.
- Barnes M, Schnitzler A, Madeiros L, et al. 2010. Efficacy and safety of NT 201 for upper limb spasticity of various etiologies - a randomized parallel-group study. *Acta Neurol Scand* 122(4):295-302.
- Borodic GE, Joseph M, Fay L, et al. 1990. Botulinum A toxin for the treatment of spasmodic torticollis: dysphagia and regional toxin spread. *Head Neck* 12(5):392-399.
- Brashear A, Gordon MF, Elovic E, et al. 2002. Intramuscular injection of botulinum toxin for the treatment of wrist and finger spasticity after a stroke. *N Engl J Med* 347(6):395-400.
- Brashear A, Lew MF, Dykstra DD, et al. 1999. Safety and efficacy of NeuroBloc (botulinum toxin type B) in type A-responsive cervical dystonia. *Neurology* 53(7):1439-1446.
- Brin MF, Comella CL, Jankovic J, et al. 2008. Long-term treatment with botulinum toxin type A in cervical dystonia has low immunogenicity by mouse protection assay. *Mov Disord* 23(10):1353-1360.
- Brin MF, Lew MF, Adler CH, et al. 1999. Safety and efficacy of NeuroBloc (botulinum toxin type B) in type A-resistant cervical dystonia. *Neurology* 53(7):1431-1438.
- Brunger AT, Breidenbach MA, Jin R, et al. 2007. Botulinum neurotoxin heavy chain belt as an intramolecular chaperone for the light chain. *PLoS Pathog* 3(9):1191-1194.
- Butler AG, Duffey PO, Hawthorne MR, et al. 2004. An epidemiologic survey of dystonia within the entire population of northeast England over the past nine years. *Adv Neurol* 94:95-99.
- Carruthers A, Carruthers J, 2005b. Said S. Dose-ranging study of botulinum toxin type A in the treatment of glabellar rhytids in females. *Dermatol Surg* 31(4):414-422.
- Carruthers A, Carruthers J, Cohen J. 2003. A prospective, double-blind, randomized, parallel-group, dose-ranging study of botulinum toxin type a in female subjects with horizontal forehead rhytides. *Dermatol Surg* 29(5):461-467.
- Carruthers A, Carruthers J. 2005a. Prospective, double-blind, randomized, parallel-group, dose-ranging study of botulinum toxin type a in men with glabellar rhytids. *Dermatol Surg* 31(10):1297-1303.
- Carruthers A, Kiene K, Carruthers J. 1996. Botulinum A exotoxin use in clinical dermatology. *J Am Acad Dermatol* 34(5 Pt 1):788-797.
- Carruthers JA, Lowe NJ, Menter MA, et al. 2002. A multicenter, double-blind, randomized, placebo-controlled study of the efficacy and safety of botulinum toxin type A in the treatment of glabellar lines. *J Am Acad Dermatol* 46(6):840-849.
- Cather JC, Menter A. 2002. Update on botulinum toxin for facial aesthetics. *Dermatol Clin* 20(4):749-761.

- Comella C, Jankovic J, Truong D, et al. 2009. Efficacy and safety of NT 201 (Botulinum neurotoxin free from complexing proteins) in cervical dystonia. AAN 2009 Scientific Abstract Listing; 80.
- Comella CL, Jankovic J, Shannon KM, et al. 2005. Comparison of botulinum toxin serotypes A and B for the treatment of cervical dystonia. *Neurology* 65(9):1423-1429.
- Comella CL, Jankovic J, Truong D, et al. 2009. Efficacy and Safety of NT 201 (Neurotoxin free from complexing proteins) in Cervical Dystonia. *Neurology* 72(Suppl 3):A345.
- Comella CL, Leurgans S, Stebbins GT. 1999. Factor Analysis and Internal Consistency of the Toronto Western Spasmodic Torticollis Rating Scale (TWSTRS). *Mov Disord* 14(5):903.
- Comella CL, Stebbins GT, Goetz CG, et al. 1997. Teaching tape for the motor section of the Toronto Western Spasmodic Torticollis Scale. *Mov Disord* 12(4):570-575.
- Comella CL. 2008; The treatment of cervical dystonia with botulinum toxins. *J Neural Transm* 115(4):579-583.
- Consky E, Basinski A, Belle L, et al. 1990. The Toronto Western Spasmodic Torticollis Rating Scale (TWSTRS): Assessment of Validity and Inter-Rater Reliability. *Neurology* 40(Suppl 1):1199P.
- Consky ES, Lang AE. 1994. Clinical Assessments of Patients with Cervical Dystonia. In: Jankovic J, Hallett M, editors. *Therapy with Botulinum Toxin*. (15):211-237. New York, Marcel Dekker.
- Costa J, Espirito-Santo C, Borges A, et al. 2004. Botulinum toxin type A therapy for blepharospasm. *The Cochrane Database of Systematic Reviews Issue 2*. Art No.: CD004900.pub2. DOI:10.1002/14651858.CD004900.pub2:
- Costa J, Espirito-Santo C, Borges A, et al. 2005a. Botulinum Toxin Type A Therapy for Cervical Dystonia. *The Cochrane Database of Systematic Reviews Issue 1*. Art No.: CD003633.pub2. DOI:10.1002/14651858.CD003633.pub2:
- DasGupta BR, Sathyamoorthy V. 1984. Purification and Amino Acid Composition of Type A Botulinum Neurotoxin. *Toxicon* 22(3):415-424.
- Dauer WT, Burke RE, Greene P, et al. 1998. Current concepts on the clinical features, aetiology and management of idiopathic cervical dystonia. *Brain* 121(Pt 4):547-560.
- De Almeida AT, de BK. 2007a. Diffusion characteristics of botulinum neurotoxin products and their clinical significance in cosmetic applications. *J Cosmet Laser Ther* 9(Suppl 1):17-22.
- De Paiva A, Meunier FA, Molgo J, et al. 1999. Functional repair of motor endplates after botulinum neurotoxin type A poisoning: Biphasic switch of synaptic activity between nerve sprouts and their parent terminals. *Proc Natl Acad Sci U S A* 96(6):3200-3205.
- Defazio G, Livrea P, De Salvia R, et al. 2001. Prevalence of primary blepharospasm in a community of Puglia region, Southern Italy. *Neurology* 56(11):1579-1581.
- Dodd SL, Rowell BA, Arrowsmith RJ, et al. 1998. A comparison of the spread of three formulations of botulinum neurotoxin A as determined by effects on muscle function. *Eur J Neurol* 5(2):181-186.
- Dolly JO. 1997. Therapeutics and research exploitation of botulinum neurotoxins. *Eur J Neurol* 4(Suppl 2):S5-S10.
- Dong M, Yeh F, Tepp WH, et al. 2006. SV2 is the protein receptor for botulinum neurotoxin A. *Science* 312(5773):592-596.
- Dressler D, Bigalke H, Rothwell JC. 2000b. The sternocleidomastoid test: An in vivo assay to investigate botulinum toxin antibody formation in humans. *J Neurol* 247(8):630-632.

- Dressler D, Dirnberger G, Bhatia KP, et al. 2000c. Botulinum Toxin Antibody Testing: Comparison Between the Mouse Protection Assay and the Mouse Lethality Assay. *Mov Disord* 15(5):973-976.
- Dressler D, Hallett M. 2006. Immunological aspects of Botox, Dysport and Myob-loc/NeuroBloc. *Eur J Neurol* 13(Suppl 1):11-15.
- Dunne JW, Heye N, Dunne SL. 1995. Treatment of chronic limb spasticity with botulinum toxin A. *J Neurol Neurosurg Psychiatry* 58(2):232-235.
- Eisele KH, Fink K, Vey M, et al. 2011. Studies on the dissociation of botulinum neurotoxin type A complexes. *Toxicon* 57(4):555-565.
- Eisele K-H, Taylor HV. 2008. Dissociation of the 900kDa neurotoxin complex from C. botulinum under physiological conditions. *Toxicon* 51(1):10. *Toxins 2008 Baveno, Italy*.
- Epidemiologic Study of Dystonia in Europe. Sex-related influences on the frequency and age of onset of primary dystonia. *Neurology* 1999;53(8):1871-1873.
- Evidente VG, Scottsdale A, Fernandez H, et al. 2011. Efficacy and Safety of Repeated Inco-botulinumtoxinA Injections in Subjects with Cervical Dystonia Previously Treated with Botulinum Toxin. *Neurology* 76(Suppl. 4):A540-A541.
- Fernandez H, Comella C, Brodsky M, et al. 2011. Efficacy and Safety of Repeated Inco-botulinum toxin A Injections in Subjects with Cervical Dystonia Previously Naive to Botulinum Toxin Treatment. *Neurology* 76(Suppl. 4):A330.
- Filippi GM, Errico P, Santarelli R, et al. 1993. Botulinum A toxin effects on rat jaw muscle spindles. *Acta Otolaryngol* 113(3):400-404.
- Frevert J. 2010. Content of botulinum neurotoxin in botox /vistabel , dysport /azzalure , and xeomin /bocouture . *Drugs R D* 10(2):67-73.
- Fujinaga Y, Inoue K, Watanabe S, et al. 1997. The haemagglutinin of Clostridium botulinum type C progenitor toxin plays an essential role in binding of toxin to the epithelial cells of guinea pig small intestine, leading to the efficient absorption of the toxin. *Microbiology* 143:3841-3847.
- Garner CG, Straube A, Witt TN, et al. 1993. Time course of distant effects of local injections of botulinum toxin. *Mov Disord* 8(1):33-37.
- Geigy Scientific Tables. Volume 3: Physical Chemistry, Composition of Blood, Hematology, Somatometric Data. 1984; 8th Edition. Basel, Ciba-Geigy AG.
- Giladi N. 1997. The mechanism of action of botulinum toxin type A in focal dystonia is most probably through its dual effect on efferent (motor) and afferent pathways at the injected site. *J Neurol Sci* 152(2):132-135.
- Göschel H, Wohlfarth K, Frevert J, et al. 1997. Botulinum A toxin therapy: Neutralizing and nonneutralizing antibodies - therapeutic consequences. *Exp Neurol* 147(1):96-102.
- Gracies JM. 2005. Pathophysiology of spastic paresis. II: Emergence of muscle overactivity. *Muscle Nerve* 31(5):552-571.
- Grafe S, Comella C, Jankovic J, et al. 2009a. Efficacy and safety of NT 201; botulinum neurotoxin type A free from complexing proteins) in treatment-naive cervical dystonia patients. *Mov Disord* 24(1):S92-S93.
- Grafe S, Comella C, Jankovic J, et al. 2009b. Efficacy and safety of NT 201, botulinum neurotoxin type A free from complexing proteins) in pre-treated cervical dystonia patients. *Mov Disord* 24(1):S92.
- Grafe S, Hanschmann A, Jankovic J. 2010. Safety and efficacy of repeated NT 201 (Botulinum neurotoxin type A free from complexing proteins) injections of patients with blepha-

- rospasm. *Mov Disord* 25(7):S235-S236. Fourteenth International Congress of Parkinson's Disease and Movement Disorders, June 13 - June 17, 2010; Buenos Aires, Argentina.
- Grandas F, Elston J, Quinn N, et al. 1988. Blepharospasm: A review of 264 patients. *J Neurol Neurosurg Psychiatry* 51(6):767-772.
- Grazko MA, Polo KB, Jabbari B. 1995. Botulinum toxin A for spasticity, muscle spasms, and rigidity. *Neurology* 45(4):712-717.
- Greene P, Fahn S, Diamond B. 1994. Development of resistance to botulinum toxin type A in patients with torticollis. *Mov Disord* 9(2):213-217.
- Greene P, Kang U, Fahn S, et al. 1990. Double-blind, placebo-controlled trial of botulinum toxin injections for the treatment of spasmodic torticollis. *Neurology* 40(8):1213-1218.
- Hagenah R, Benecke R, Wiegand H. 1977. Effects of type A botulinum toxin on the cholinergic transmission at spinal Renshaw cells and on the inhibitory action at Ia inhibitory interneurons. *Naunyn Schmiedebergs Arch Pharmacol* 299(3):267-272.
- Hallett M. 1999. One man's poison - clinical applications of botulinum toxin. *N Engl J Med* 341(2):118-120.
- Hatheway CL, Dang C. 1994. Immunogenicity of the Neurotoxins of *Clostridium botulinum*. In: Jankovic J, Hallett M, editors. *Therapy with Botulinum Toxin*. (8):93-107. New York, Marcel Dekker.
- Herrera AA. 1997. Remodeling and Elimination of Neuromuscular Synapses. *The Synapse in Development, Health, and Disease. Advances in Organ Biology* (2):29-53. Greenwich CT, JAI Press Inc. Festoff BW, Hantai D, Citron B.
- Jankovic J, Comella C, Hanschmann A, et al. 2011b. Efficacy and safety of incobotulinumtoxinA (NT 201, Xeomin) in the treatment of blepharospasm-A randomized trial. *Mov Disord*; In Press: 10.1002/mds.23658 (doi)
- Jankovic J, Kenney C, Grafe S, et al. 2009b. Relationship Between Various Clinical Outcome Assessments in Patients with Blepharospasm. *Mov Disord* 24(3):407-413.
- Jankovic J, Orman J. 1984. Blepharospasm: Demographic and clinical survey of 250 patients. *Ann Ophthalmol* 16(4):371-376.
- Jankovic J, Orman J. 1987. Botulinum A toxin for cranial-cervical dystonia: A double-blind, placebo-controlled study. *Neurology* 37(4):616-623.
- Jankovic J. 2004b. Botulinum toxin in clinical practice. *J Neurol Neurosurg Psychiatry* 75(7):951-957.
- Jost WH, Blümel J, Grafe S. 2007. Botulinum Neurotoxin Type A Free of Complexing Proteins (Xeomin) in Focal Dystonia. *Drugs* 67(5):669-683.
- Jost WH, Kohl A, Brinkmann S, et al. 2005. Efficacy and tolerability of a botulinum toxin type A free of complexing proteins (NT201) compared with commercially available botulinum toxin type A (BOTOX) in healthy volunteers. *J Neural Transm* 112(7):905-913.
- Kanovsky P, Slawek J, Denes Z, et al. 2009a. Efficacy and Safety of Botulinum Neurotoxin NT 201 in Poststroke Upper Limb Spasticity. *Clin Neuropharmacol* 32(5):259-265.
- Kanovsky P, Slawek J, Denes Z, et al. 2011. Efficacy and safety of treatment with Inco-botulinum toxin A (botulinum neurotoxin type A free from complexing proteins; NT 201) in post-stroke upper limb spasticity. *J Rehabil Med* 43(6):486-492.
- Kessler KR, Benecke R. 1997. Botulinum toxin: From poison to remedy. *Neurotoxicology* 18(3):761-770.

- Lance JW. Symposium Synopsis. In: Feldman RG, Young RR, Koella WP, editors. 1980. rigidity, disordered motor control. 485-494. Miami, Symposia Specialists.
- Lew MF, Adornato BT, Duane DD, et al. 1997. Botulinum toxin type B: A double-blind, placebo-controlled, safety and efficacy study in cervical dystonia. *Neurology* 49(3):701-707.
- Lu CS, Chen RS, Tsai CH. 1995. Double-blind, placebo-controlled study of botulinum toxin injections in the treatment of cervical dystonia. *J Formos Med Assoc* 94(4):189-192.
- Mahrhold S, Rummel A, Bigalke H, et al. 2006. The synaptic vesicle protein 2C mediates the uptake of botulinum neurotoxin A into phrenic nerves. *FEBS Lett* 580(8):2011-2014.
- Marras C, Van den Eeden SK, Fross RD, et al. 2007. Minimum incidence of primary cervical dystonia in a multiethnic health care population. *Neurology* 69(7):676-680.
- Marx M, Grafe S, Comella C. 2010. Use of an interactive voice response system by patients with blepharospasm receiving repeated injections of NT 201 (botulinum neurotoxin Type A free from complexing proteins). *Mov Disord* 25(7):S243. Fourteenth International Congress of Parkinson's Disease and Movement Disorders, June 13 - June 17, 2010; Buenos Aires, Argentina.
- Mauriello JA, Dhillon S, Leone T, et al. 1996b. Treatment selections of 239 patients with blepharospasm and Meige syndrome over 11 years. *Br J Ophthalmol* 80(12):1073-1076.
- Minor V. 1928. Ein neues Verfahren zu der klinischen Untersuchung der Schweißabsonderung. *Deutsch Z Nervenheilk* 101:302-306.
- Monheit G, Carruthers A, Brandt F, et al. 2007. A randomized, double-blind, placebo-controlled study of botulinum toxin type A for the treatment of glabellar lines: determination of optimal dose. *Dermatol Surg* 33(1 Spec No.):S51-S59.
- Nakashima K, Kusumi M, Inoue Y, et al. 1995. Prevalence of focal dystonias in the western area of Tottori Prefecture in Japan. *Mov Disord* 10(4):440-443.
- Naumann M, Yakovlev A, Durif F. 2002. A randomized, double-masked, crossover comparison of the efficacy and safety of botulinum toxin type A produced from the original bulk toxin source and current bulk toxin source for the treatment of cervical dystonia. *J Neurol* 249(1):57-63.
- Pappert EJ, Germanson T. 2008. Botulinum toxin type B versus type A in toxin-naive patients with cervical dystonia: Randomized, double-blind, noninferiority trial. *Mov Disord* 23(4):510-517.
- Pellizzari R, Rossetto O, Schiavo G, et al. 1999. Tetanus and botulinum neurotoxins: Mechanism of action and therapeutic uses. *Philos Trans R Soc Lond B Biol Sci* 354(1381):259-268.
- Pickett A, Dodd S, Rzany B. 2008. Confusion about diffusion and the art of misinterpreting data when comparing different botulinum toxins used in aesthetic applications. *J Cosmet Laser Ther* 181-183.
- Roggenkämper P, Jost WH, Bihari K, et al. 2006. Efficacy and safety of a new Botulinum Toxin Type A free of complexing proteins in the treatment of blepharospasm. *J Neural Transm* 113(3):303-312.
- Rosales RL, Arimura K, Takenaga S, et al. 1996. Extrafusal and intrafusal muscle effects in experimental botulinum Neurotoxin-A injection. *Muscle Nerve* 19(4):488-496.
- Rzany B, Ascher B, Fratila A, et al. 2006. Efficacy and safety of 3- and 5-injection patterns (30 and 50 U) of botulinum toxin A (Dysport) for the treatment of wrinkles in the glabella and the central forehead region. *Arch Dermatol* 142(3):320-326.

- Sattler G, Callander MJ, Grablowitz D, et al. 2010. Noninferiority of incobotulinumtoxinA, free from complexing proteins, compared with another botulinum toxin type A in the treatment of glabellar frown lines. *Dermatol Surg* 36(Suppl 4):2146-2154.
- Schiavo G, Montecucco C. 1997. The Structure and Mode of Action of Botulinum and Tetanus Toxins. In: *The Clostridia: Molecular Biology and Pathogenesis*. (18):295-322. Academic Press Limited.
- Schnider P, Brichta A, Schmied M, et al. 1993. Gallbladder dysfunction induced by botulinum A toxin. *Lancet* 342(8874):811-812.
- Sharma SK, Singh BR. 2000. Immunological properties of HN-33 purified from type A *Clostridium botulinum*. *J Nat Toxins* 9(4):357-362.
- Simpson DM, Alexander DN, O'Brien CF, et al. 1996. Botulinum toxin type A in the treatment of upper extremity spasticity: a randomized, double-blind, placebo-controlled trial. *Neurology* 46(5):1306-1310.
- Simpson DM, Blitzer A, Brashear A, et al. 2008a. Assessment: Botulinum neurotoxin for the treatment of movement disorders (an evidence-based review): report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology* 70(19):1699-1706.
- Simpson DM, Gracies JM, Graham HK, et al. 2008b. Assessment: Botulinum neurotoxin for the treatment of spasticity (an evidence-based review): report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology*;70(19):1691-1698.
- Sloop RR, Cole BA, Escutin RO. 1997. Human response to botulinum toxin injection: Type B compared with type A. *Neurology* 49(1):189-194.
- Sommerfeld DK, Eek EU, Svensson AK, et al. Spasticity after stroke: its occurrence and association with motor impairments and activity limitations. *Stroke* 35(1):134-139.
- Tang-Liu DD, Aoki KR, Dolly JO, et al. 2003. Intramuscular injection of 125I-botulinum neurotoxin-complex versus 125I-botulinum-free neurotoxin: time course of tissue distribution. *Toxicon* 42(5):461-469.
- Thussu A, Barman CR, Prabhakar S. 1999. Botulinum toxin treatment of hemifacial spasm and blepharospasm: Objective response evaluation. *Neurol India* 47(3):206-209.
- Truong D, Comella C, Fernandez HH, et al. 2008a. Efficacy and safety of purified botulinum toxin type A (Dysport) for the treatment of benign essential blepharospasm: A randomized, placebo-controlled, Phase II trial. *Parkinsonism Relat Disord* 407-414.
- Truong D, Lew M, Orlova O. 2008. Efficacy of Dysport (Botulinum toxin type A) for the treatment of cervical dystonia: A multicenter, randomized, double-blind, placebo-controlled Phase III study. *Toxicon* 51(1):37-38. *Toxins 2008 Baveno, Italy*.
- Tsui JK, Eisen A, Stoessl AJ, et al. 1986. Double-blind study of botulinum toxin in spasmodic torticollis. *Lancet* 2(8501):245-247.
- Ward AB, Aguilar M, De Beyl Z, et al. 2003. Use of botulinum toxin type A in management of adult spasticity - A European consensus statement. *J Rehabil Med* 35(2):98-99.
- Watkins CL, Leathley MJ, Gregson JM, et al. 2002. Prevalence of spasticity post stroke. *Clin Rehabil* 16(5):515-522.
- WE MOVE. Management of Spasticity with Botulinum Toxin Type A (Botox). 2005b; (3.0) <http://www.mdvu.org>

- Wiegand H, Erdmann G, Wellhöner HH. 1976. 125I-labelled botulinum A neurotoxin: Pharmacokinetics in cats after intramuscular injection. *Naunyn Schmiedebergs Arch Pharmacol* 292(2):161-165.
- Wissel J, Müller J, Dressnandt J, et al. 2000. Management of spasticity associated pain with botulinum toxin A. *J Pain Symptom Manage* 20(1):44-49.
- Wohlfarth K, Müller C, Sassin I, et al. 2007. Neurophysiological double-blind trial of a botulinum neurotoxin type A free of complexing proteins. *Clin Neuropharmacol* 30(2):86-94.

Therapeutic Goods Administration

PO Box 100 Woden ACT 2606 Australia

Email: info@tga.gov.au Phone: 1800 020 653 Fax: 02 6232 8605

<http://www.tga.gov.au>