Supplementary data file

Appendix 1. Outcome measures

Outcome measures

Disability was measured using the AMC Linear Disability Score (ALDS) (a generic self-assessment scale which includes instrumental activities of daily living (ADL).[1] The ALDS ranges from 0 and 100; higher scores indicate a better functional status.

Quality of life (QoL) was assessed using the 36-item Short Form Health Survey (SF-36).[2] Two dimension scores can be derived from 8 subscale scores: the Physical Component Summary and the Mental Component Summary. Both summary scores were normalized to a general Dutch population mean of 50 and a standard deviation of 10. [3] Higher scores indicate a better HRQL.[4]

Psychiatric evaluation included the Mini-International-Neuropsychiatric Interview – Plus (MINI-Plus), screening for possible psychiatric co-morbidity according to DSM-IV criteria. [4]

Several quantitative questionnaires concerning psychiatric symptoms were assessed too, including the Montgomery Asberg Depression Rating Scale (MADRS) (10 item interview, scores ranging from 0-60, higher scores indicating more severe symptoms) and The Beck Depression Inventory (BDI)[5] (21 item self-report scale, ranging from 0-63, higher scores indicating more severe symptoms) for depressive symptoms; the Beck Anxiety Inventory (BAI)[6] (21 item self-report scale, ranging from 0-63, higher scores indicating more severe symptoms) and Liebowitz Social Anxiety Scale (LSAS)[7] (48 item self-report scale, scores ranging from 0-144, higher scores indicating more severe symptoms) for anxiety and the Obsessive-Compulsive Inventory (OCI-R)[8] (18 item self-report, scores ranging from 0-72, higher scores indicating more severe symptoms) for obsessive-compulsive symptoms.

References

1. Holman R, Weisscher N, Glas CA, et al. The Academic Medical Center Linear Disability Score (ALDS) item bank: item response theory analysis in a mixed patient population. *Health QualLife Outcomes* 2005; **3**: 83.

2. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992; **30**(6): 473-83.

3. Aaronson NK, Muller M, Cohen PD, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol* 1998; **51**(11): 1055-68.

4. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998; **59 Suppl 20**: 22-33;quiz 4-57.

5. Beck AT, Steer RA. Internal consistencies of the original and revised Beck Depression Inventory. *J Clin Psychol* 1984; **40**(6): 1365-7.

6. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. *J Consult Clin Psychol* 1988; **56**(6): 893-7.

7. Heimberg RG, Horner KJ, Juster HR, et al. Psychometric properties of the Liebowitz Social Anxiety Scale. *Psychol Med* 1999; **29**(1): 199-212.

8. Foa EB, Huppert JD, Leiberg S, et al. The Obsessive-Compulsive Inventory: development and validation of a short version. *Psychol Assess* 2002; **14**(4): 485-96.

Appendix 2. Additional information on methods and results

Methods

Procedure

Patients were told that BoNT is a common treatment in movement disorders, but that it's efficacy is unknown in functional jerks/tremor. Tot test this half of the patients would receive BoNT and the other half placebo, which was sterile saline. After two treatment sessions, they would be enrolled in the open-label study where everybody would receive BoNT. Then we would explain how the injections would take place and what common side effects of BoNT are. Also they were instructed when to contact us (hematoma, severe weakness, severe adverse event).

Inter-observer analysis

Average weighted Kappa and ICC values of pairs of observations were considered as an overall index for concordance among observers. [1] Kappa and ICC values were arbitrarily classified according to Landis and Koch [2] with values <0 indicating no agreement, 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement.

Results

Screening

Of all excluded patients, 109 patients did not fulfill the inclusion criteria. In 52 (47.8%) of the 109 patients symptoms diminished severely or resolved. In the other 57 (52,3%) reasons were diverse including: not functional (n=13), no jerks/tremor or jerks/tremor not amendable for injection (n=28), terminally ill (n=5), insufficient knowledge of Dutch language (n=2), too old or too young (n=2), previous treatment with BonT without effect (n=1), coagulation disorder (n=1), too much previous therapies (n=2), moving to other country (n=1), one amputated arm (n=1), complaints present < one

year (n=1). The 18 patients who were excluded for 'other reasons' included: could not be reached (n=16), not approached because files legal complaints (n=1), death wish and died in 2012 (n=1). We don't have follow-up data on these patients unfortunately.

Selection muscles and doses used per subject

Subject	Randomisation	Axial or extremity	Muscles injected	Unilateral or bilateral	Starting dose per muscle
Pin 1	placebo	extremity	• pectoral muscle	• unilateral	100
Pin 2	botulinum	axial	• iliopsoas muscle	• bilateral	200
Pin 3	botulinum	axial	• iliopsoas muscle	• bilateral	200
Pin 4	botulinum	extremity	• trapezius muscle • teres major muscle • major pectoral muscle	• bilateral • bilateral • bilateral	50 50 50
Pin 5	placebo	axial	• paraspinal muscle Th12 • paraspinal muscle L2	• unilateral • unilateral	100 100
Pin 6	placebo	extremity	• SCM • levator scapulae muscle • trapezius muscle	• unilateral • unilateral • unilateral	60 60 120
Pin 7	placebo	axial	• iliopsoas muscle	• bilateral	200
Pin 8	botulinum	axial	• trapezius muscle • levator scapulae muscle • rectus abdominis muscle	• bilateral • bilateral • bilateral	60 60 120
Pin 9	placebo	axial	• iliopsoas muscle	• bilateral	200
Pin 10	botulinum	axial	• iliopsoas muscle	• unilateral	200
Pin 11	botulinum	extremity	• iliopsoas muscle • rectus femoris muscle • vastus medialis muscle	• unilateral • unilateral • unilateral	200 100 50
Pin 12	placebo	axial	• rectus abdomonis muscle	• bilateral	120

Pin 13	botulinum	extremity	• trapezius muscle • levator scapulae muscle	• unilateral • unilateral	80 80
Pin 14	placebo	extremity	• SCM muscle • trapezius muscle • pectoral muscle • deltoid muscle	• bilateral • bilateral • bilateral • bilateral	80 80 80 80
Pin 15	placebo	extremity	• trapezius muscle • major pectoral muscle	• bilateral • bilateral	50 100
Pin 16	placebo	axial	• paraspinal muscle Th8	• bilateral	150
Pin 17	botulinum	axial	• iliopsoas muscle • rectus abdominis muscle	• bilateral • bilateral	160 120
Pin 18	placebo	axial	• iliopsoas muscle • rectus femoris muscle	• unilateral • unilateral	200 200
Pin 19	botulinum	axial	• rectus abdominal muscle	• bilateral	200
Pin 20	botulinum	axial	• semispinal muscle • rectus abdominis muscle	• bilateral • bilateral	60 120
Pin 21	botulinum	extremity	• vastus medialis muscle • rectus femoris muscle	• bilateral • bilateral	100 100
Pin 22	placebo	axial	• rectus abdominis muscle	• bilateral	150
Pin 23	botulinum	extremity	• SCM muscle	• bilateral	40
Pin 24	placebo	axial	• iliopsoas muscle • rectus abdominis muscle	• bilateral • bilateral	120 120
Pin 25	placebo	extremity	• major pectoral muscle	• bilateral	60

Pin 26	placebo	extremity	• trapezius muscle • major pectoral muscle • deltoid muscle	• unilateral • unilateral • unilateral	40 80 80
Pin 27	botulinum	axial	• rectus abdominis muscle • iliopsoas muscle • vastus medialis muscle	• unilateral • unilateral • unilateral	120 120 60
Pin 28	botulinum	axial	• rectus abdominis muscle	• bilateral	120
Pin 29	botulinum	axial	• iliopsoas muscle	• bilateral	200
Pin 30	botulinum	extremity	• frontal muscle • auricularis superior muscle	• unilateral • unilateral	10 10
Pin 31	placebo	extremity	• biceps brachii muscle • deltoid muscle • extensor carpi radial muscle	• unilateral • unilateral • unilateral	120 120 40
Pin 32	botulinum	extremity	• pectoral muscle	• bilateral	80
Pin 33	botulinum	extremity	• biceps brachii muscle • flexor carpi radial muscle	• unilateral • unilateral	100 60
Pin 34	placebo	extremity	• flexor carpi radial muscle • extensor carpi radial muscle • pronator teres muscle	• unilateral • unilateral • unilateral	80 30 40
Pin 35	placebo	extremity	• abductor digiti V muscle	• unilateral	30
Pin 36	botulinum	extremity	• flexor carpi radial muscle • extensor carpi radial	• unilateral • unilateral	120 40

			muscle		
Pin 37	placebo	axial	• rectus abdominis muscle • iliopsoas muscle	• bilateral • bilateral	120 100
Pin 38	placebo	extremity	• splenius capitis muscle • SCM muscle	• unilateral • unilateral	100 30
Pin 39	botulinum	extremity	• supinator teres muscle • pronator teres muscle	• unilateral • unilateral	40 80
Pin 40	placebo	axial	• rectus abdominis muscle • abdominal oblique muscle	• bilateral • bilateral	120 60
Pin 41	placebo	extremity	• extensor carpi radial muscle • flexor carpi radial muscle • triceps brachii muscle	• unilateral • unilateral • unilateral	40 80 80
Pin 42	botulinum	extremity	• quadriceps femoris muscle	• unilateral	160
Pin 43	placebo	extremity	• extensor carpi radial muscle • triceps brachii muscle	• unilateral • unilateral	40 120
Pin 44	botulinum	axial	• rectus abdominis muscle	• bilateral	120
Pin 45	placebo	axial	• paraspinal muscle Th12-L5	• bilateral	180
Pin 46	botulinum	extremity	• pterygoideus lateral muscle • depressor anguli oris muscle • platysma muscle	• unilateral • unilateral • unilateral	30 10 10

Pin 47	botulinum	axial	• major pectoral muscle	• bilateral	60
Pin 48	botulinum	axial	• iliopsoas muscle	• bilateral	160

Open-label extension

Compared to the end of the trial 19 of 43 patients (44.2%) showed improvement (score 1,2 or 3) of motor symptoms on the CGI-I assessed by the investigators (score 1 n=5 (12.0%); score 2 n=5 (11.0%); score 3 n=9 (20.9%); score 4 n=19 (44.2%); score 5 n=3 (7.0%); score 6 n=2 (4.7%); score 7 n=0) (**see figure 3**). Compared to baseline motor symptom improvement (score 1,2 or 3) occurred in 35 of 43 patients (81.4%) (score 1 n=10 (23.3%); score 2 n=10 (23.3 %); score 3 n=15 (34.9%); score 4 n=5 (11.6%); score 5 n=3 (7.0%), score 6 or 7 n=0).

The CGI-I scored by the patient revealed a perceived motor improvement compared to the end of trial in 24 of 43 patients (55.8%) (score 1 n=5 (11.5%); score 2 n=11 (25.0%); score 3 n=8 (18.2%); score 4 n=17 (38.6%); score 5 n=1 (2.3%); score 6 n=1 (2.3%); score 7 n=1 (2.3%)) (**see figure 3**). Compared to baseline perceived motor symptom improvement occurred in 29 of 43 patients (67.4%) (score 1 n=11 (25.6%); score 2 n=14 (32.6%); score 3 n=4 (9.3%); score 4 n=12 (27.9%); score 5 n=1 (2.3%); score 6 n=6 (2.3%); score 7 n=0).

References

1. Feinstein A. Clinimetric Perspectives. J Chron Dis 1987;40:635-40.

2. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

Appendix 3. Statistical analysis plan

Overall principles

The data analysis will start after the 12-month follow-up data of the last included patient has been obtained, and after the study database has been cleaned and locked.

The analyses will be done by investigator (YD) supervised by the principal investigator (MAJdKT) and an independent epidemiologist/statistician of the AMC Clinical Research Unit (RdH). The statistical programming and analysis to produce all summary tables and figures will use the statistical package IBM SPSS statistics version 22.

In general, variables will be summarized using simple descriptive statistics such as means with standard deviation for continuous symmetric variables, medians and interquartile ranges for continuous skewed variables, and frequencies with percentages for categorical variables.

All analyses will be done according to the intention-to-treat principle, by analysing patients in the groups to which they were allocated by randomisation. The analyses will first be performed blind to treatment allocation, to allow for checking of the data and the proposed summaries/analyses. After the investigation and correction of any isolated or systematic data errors, treatment allocation will be unmasked.

The primary outcome will be analysed in the pre-specified subgroups below, irrespective of the presence of statistical significance in the overall analysis. Safety outcomes will be additionally analysed in the as-treated (not per-protocol) population.

Overall level of statistical significance

According to Haybittle-Peto's stopping rule, no adjustment of the p-value will be used for the final analysis [20]. A two-sided p-value < 0.05 will be considered statistically significant. Statistical uncertainty will be expressed in a two-sided 95% CI.

Missing data

Missing baseline and outcome data will not be imputed. We will state which data are missing and calculate frequencies using the total number of patients with available data. When a patient is lost to follow-up or has withdrawn consent, we will use all available data up until withdrawal of consent or loss to follow-up. A specific section in the paper will report on missing data.

Poweranalysis

A two group Chi-square test with a 0,05 two-sided significance level will have 80% power to detect the difference between a control group proportion of 0,30 and a treatment group proportion of 0,70 (odds ratio of 5,4) when the sample size in each group is 24. As the side effects of therapy are mild and self-limiting and because only two injections are given in the trial period, we expect practically no withdrawals in this phase of the study. Assuming a withdrawal rate of 10 percent, we plan to include 27 patients per treatment arm, which means 54 patients in total.

Population

Intention-to-treat population

All randomised patients will be analysed in the treatment group to which they were originally allocated irrespective of non-adherence or deviations from protocol.

As-treated population

Patients will be analysed in groups according to treatment received. The patients will still be included in the as-treated analysis if there was a protocol violation (e.g. not receiving treatment within the described time-frame, not receiving the correct treatment, or not meeting inclusion or exclusion criteria).

List of analyses

Recruitment and retention

The trial profile and inclusion will be shown in a CONSORT flow diagram *(figure 1),* including the total number of randomised patients and then showing per treatment group the numbers receiving allocated treatment, withdrawing consent, and lost to follow up.

Baseline characteristics

Table 1: Patient characteristics

Treatment group vs placebo group

Age (mean of median)

Gender

Duration of symptoms (mean or median)

Site of jerks

Abdominal vs extremity

Medication usage

Presence of bereitschaftspotential

Protocol deviations and violations

All substantial protocol violations will be listed.

Adherence to allocated treatment

Adherence will be reported descriptively.

Primary outcome

An intention to treat analysis will be performed with regard to the trial results. The difference in the proportions of patients reaching the primary outcome measure (score 1,2 of 3 on CGI) between the groups treated with BoNT and with placebo will be assessed using the $\chi 2$ statistic or Fisher's exact test, when appropriate. The CGI will be dichotomized to improvement (score 1,2 of 3) vs no change or worsening (score 4,5,6,7). A binary logistic regression analysis will be performed to correct for the treatment group (BoNT vs. placebo) and stratification-factor (axialvs. extremity). The effect size will be expressed in an odd's ratio.

Secondary outcomes

For the secondary outcome measures, proportional differences between the groups will be tested with the $\chi 2$ statistic or Fischer's exact test when appropriate. Difference in change scores of the continuous secondary outcome measures will be calculated. The mean or median differences will be analysed with a Students t-test or Mann-Whitney (when appropriate). Statistical uncertainty will be expressed with a 95% CI.

For the long term effects, the within group change scores at t=16 months will be compared with previous assessments (end of trial and baseline) using tests for paired data (Wilcoxon Signed Rank).

Safety outcomes

Safety outcomes will be reported in the intention-to treat and as-treated populations using descriptive statistics.

Subgroup analyses

No subgroup analysis will be performed because of the small amount of patients and hence the lack of power.

Appendix 4. Video protocol

1. Informed consent

2. Rest (sitting on research bench)

Total body	2 minutes
Focus on face	20 seconds
Reading standard text	

3. Action

Stretching out arms, palms facing downwards	30 seconds
Stretching out arms, palms facing upwards	30 seconds
Bending arms in front of chest	20 seconds
Fingertapping	10x
Bradykinesia	10x
Finger-to-nose (right and left)	5x
Stretching out leg (right and left)	20 sec

Attention task	Subtracting 100 minus 7
	Subtracting 100 minus 13
	Counting the months of the year backwards
Entrainment:	tapping along with metronome: 112 en 138 bpm

Suppressing symptoms	10 seconds

4. Leaving patient alone in the room	2 minutes
5. Standing	
From four different angles	30 seconds
6. Research bench	
- Testing tendon reflexes	
- Lying supine	30 seconds
7. Gait	30 seconds