Treating Depression with Botulinum Toxin: A Pooled Analysis of Randomized Controlled Trials

Authors

M. Magid1 *, E. Finzi2 *, T. H. C. Kruger1 *, H. T. Robertson1, B. H. Keeling1, S. Jung1, 8, J. S. Reichenberg5, N. E. Rosenthal6, 7, M. A. Wollmer8, 9

Affiliations

Affiliation addresses and name abbreviations (initials) are listed at the end of the article

Abstract

Introduction: Botulinum toxin A (BTA) injection into the glabellar region is currently being studied as a treatment for major depressive disorder (MDD). Here we explore efficacy data of this novel approach in a pooled analysis.

Methods: A literature search revealed 3 RCTs on this topic. Individual patient data and clinical endpoints shared by these 3 trials were pooled and analyzed as one study (n = 134) using multiple regression models with random effects.

Results: In the pooled sample, the BTA (n = 59) and the placebo group (n = 75) did not differ in the baseline variables. Efficacy outcomes revealed BTA superiority over placebo: Improvement in the Hamilton Depression Rating Scale or Montgomery-Asberg Depression Rating Scale 6 weeks after baseline was 45.7 % for BTA vs. 14.6 % for placebo (p < 0.0001), corresponding to a BTA response rate of 54.2 % (vs. 10.7 %) and a BTA remission rate of 30.5 % (vs. 6.7 %).

Discussion: Equalling the status of a meta-analysis, this study increases evidence that a single treatment of BTA into the glabellar region can reduce symptoms of MDD. Further studies are needed to better understand how BTA exerts its mood-lifting effect.

Supporting Information for this article is available online at http://www.thieme-connect.de/products

* Equal contribution.

Introduction

Affecting more than 350 million individuals worldwide, depression is one of the greatest medical challenges of our time [1]. There are several effective psychotherapeutic, pharmacological, and somatic options for the treatment of depression, but still a considerable proportion of patients do not attain remission. Thus, there is a need to develop further methods of treatment.

The injection of botulinum toxin A (BTA) into the glabellar frown muscles of the forehead is an emerging novel approach in the treatment of depression: after an auspicious first case series, we recently conducted 3 independent randomized, controlled trials and consistently showed that a single glabellar treatment with BTA reduces symptoms of depression [2–5].

The treatment of glabellar frown lines for cosmetic reasons is an approved indication for BTA and is the most frequent intervention in aesthetic medicine [6]. Several studies have shown that this particular cosmetic treatment may exert psychological effects, which may contribute to its popularity [7–11]. In one study, BTA in the forehead was compared to other cosmetic treatments (i.e., BTA in other regions besides the forehead, Restylane, glycolic peels, and laser treatments) [9]. Those who received BTA in the forehead reported feeling less depressed, irritable, and anxious, as compared to the control group, despite both groups feeling “equally attractive” after the intervention. This indicates that BTA in the forehead has positive mood effects above and beyond the euphoria that one gets from improved appearance.

The mood-lifting effect of such a BTA treatment may be explained by the facial feedback hypothesis, which dates back to Charles Darwin and William James in the 19th century and has been substantiated in several experimental studies [12–14]. In theory, contraction of facial muscles sends a message to the emotional centers of the brain. Smiling can reinforce and maintain feelings of well-being, whereas frowning can lead to the opposite. We assume that the paralysis of the injected frown muscles interrupts a proprioceptive feedback loop from the face to the emotional brain, therefore reducing the ability to feel negative emotions [15–17]. In depression, there is a
relative over-activity of frown muscles, which has been confirmed electrophysiologically, and whose elimination by BTA treatment could potentially soften the correspondent experience of fear, anger, and sadness [18, 19]. In order to further corroborate the evidence of a role for BTA in the treatment of depression, we conducted a conjoint analysis of pooled individual patient data from our 3 trials.

Methods

Studies
The 3 studies from which we pooled data were all investigator-initiated, randomized controlled trials that were carried out free of any commercial entity. The studies were initiated and conducted independently from each other, however MM et al. adapted the protocol of MAW et al. to facilitate their comparison. All studies were approved by the responsible regulatory authorities and ethic committees. All patients gave written informed consent to take part in the study. All studies are registered at ClinicalTrials.gov (NCT00934687, NCT01556971, NCT01392963). Detailed descriptions of the studies are provided in the respective original articles [3–5] and key points are highlighted in Supplemental Table 1.

In summary all studies included female and male adult patients suffering from unipolar major depressive disorder (DSM-IV 296.2x and 296.3x), who were randomized to receive injections of either a total of 29 U (women) or 39 or 40 U (men) of BTA (onabotulinumtoxinA, Botox®, Allergan) or a placebo of 0.9% NaCl saline at 5 points into the corrugator supercilii and procerus muscles in the glabellar region and were assessed for change in the symptoms of depression 6 weeks thereafter. By means of a systematic literature search in PubMed, Web of Science, the Cochrane Controlled Trials Register using “[depression OR major depressive disorder] AND [botulinum toxin A OR botulinum toxin type-A OR Botox]” as key words, and Medical Subject Headings (MeSH) no other comparable studies (i.e., randomized placebo-controlled study of BTA as a treatment for depression) have been published up to December 2014.

Inclusion criteria, scales, and data pooling
For our conjoint analysis we included baseline variables and outcome measures that were shared by all 3 studies (Table 1). These variables were markedly similar throughout all 3 studies, allowing for pooling. Of note the BDI score was 4 points higher (p=0.009) in the study by EF and NER than in the other studies; the CSS-GFL was lower (p<0.0001) in this study than in the other studies (Supplemental Table 1). In the pooled sample (n=134), the BTA (n=59) and the placebo (n=75) did not differ in any of the shared baseline variables, including BDI score and CSF-GFL scores (Table 1). These 2 variables, along with other baseline variables, were accounted for using random effects models (see statistics below), presenting evidence against confounding.

In MAW et al. and MM et al. studies, patients needed to produce “moderate to severe frown lines” in order to be included. This was not an inclusion criteria in EF’s study. Of note, very few people (n=5) were excluded on this basis, as most people, except for very young adults, are able to produce at least moderate frown lines at maximum frowning.

All 3 trials studied BTA as an adjunctive treatment to antidepressant medications and particularly the study by EF also BTA as a...
monotherapy (patients not taking antidepressants) (Supplemental Table 1). Pooled data analysis allowed for comparison of BTA as an adjunct vs. a monotherapy. Tolerability and side effect data were reported in the studies done by EF et al. and MAW et al., and were collected from raw data by MM et al.

The primary end point was a reduction in depressive symptoms 6 weeks after the baseline. This was measured by the Hamilton Depression Rating Scale (HAM-D) in MAW et al. and MM et al. studies and by the Montgomery-Asberg Depression Rating Scale (MADRS) in EF et al. study. We used continuous (% change in score) and categorical (response, ≥50% improvement from baseline scores; remission, score ≤7 for the HAM-D scales, ≤10 for the MADRS) changes on these scales as common outcome variables, thus eliminating confounding factors that may have arisen from using different rating scales. Specifically, the cut-off values for remission allow for optimal comparability of the HAM-D and the MADRS scale [20]. In the data from the study by MM et al., we recalculated response and remission rates based on the 17-item version of the HAM-D instead of the originally used 21-item version to improve comparability with the study by MAW et al. All 3 studies used the Beck Depression Inventory (BDI) as a self-rating scale (remission ≤9). One of the trials used continuous (% change in score) and categorical (response, ≥50% improvement from baseline scores; remission, score ≤7 for the HAM-D scales, ≤10 for the MADRS) changes on these scales as common outcome variables, thus eliminating confounding factors that may have arisen from using different rating scales. Specifically, the cut-off values for remission allow for optimal comparability of the HAM-D and the MADRS scale [20]. In the data from the study by MM et al., we recalculated response and remission rates based on the 17-item version of the HAM-D instead of the originally used 21-item version to improve comparability with the study by MAW et al. All 3 studies used the Beck Depression Inventory (BDI) as a self-rating scale (remission ≤9). One of the trials used the newer BDI-II version. Since the 2 versions are highly correlated (r = 0.82–0.94) and have an identical point range, we used them as one scale (BDI) [21]. All studies used the same Clinical Severity Score for Glabellar Frown Lines (CSS-GFL) to assess frown scores at baseline and week 6 [22]. One of the trials involved a cross-over of the BTA and the placebo group after 12 weeks. The crossover data was not included in our conjoint analysis.

Statistical analysis

Multiple regression models incorporating random effects for each study were evaluated for the following outcomes: 1) BDI score change, 2) BDI % change, 3) BDI response rate, 4) BDI remission rate, 5) HAM-D/MADRS % change, 6) HAM-D/MADRS response rate, 7) HAM-D/MADRS remission rate, 8) CSS-GFL score change. Models 1, 2, 5 and 8 were calculated using linear mixed models. The remaining were calculated as logistic mixed models. Each model was adjusted for age, sex, baseline CSS-GFL scores, and baseline depression severity for each of the 3 studies. Other variables such as the use of antidepressants were fitted and found to be non-significant in all cases. Possible interactions between the predictor variables were also tested; none were found to be significant. Odds ratios (OR) and numbers needed to treat (NNT) were calculated based on adjustments for age, sex, and baseline depression severity. All models found that patients who received the BTA intervention had significantly better outcomes, with p-values below 0.001; the effect of the intervention was invariant and no significant interactions were found with respect to baseline score, age, and sex. Random effect models compensated for any unobserved differences between the 3 studies.

Results

BTA was superior to placebo in all psychopathological efficacy outcomes (p < 0.0001–0.03, Table 1): As for the primary end point, i.e., the improvement in the HAM-D or MADRS expert rating scale, recipients of BTA experienced an overall score reduction of 45.7% vs. 14.6% in the placebo group. This corresponds to a response rate of 54.2% (vs. 10.7%, OR 11.1, 95% CI 4.3–28.8, NNT = 2.3) and a remission rate of 30.5% (vs. 6.7%, OR 7.3, 95% CI 2.4–22.5, NNT = 4.2) (Fig. 1a). On the BDI, the BTA group improved by 14.3 points (47.4%) compared to 5.1 points (16.2%; Cohen’s d = 1.07, Fig. 2), corresponding to a response rate of 52.5% (vs. 8.0%, OR 11.1, 95% CI 4.3–28.8, NNT = 2.2) and a remission rate of 42.4% (vs. 8.0%, OR 15.7, 95% CI 4.8–50.9, NNT = 2.9) (Fig. 1b).

The response rates did not differ significantly between patients receiving BTA as a monotherapy vs. patients receiving it as an adjunctive treatment in addition to an established treatment with psychotropic medications (BDI: p = 0.61) (Supplemental Table 2).

In the pooled analysis, approximately 9.3% (n = 7) of the placebo group and 13.6% (n = 8) of the BTA reported temporary headaches after the intervention. No severe adverse reactions were
There was no statistical difference in adverse events between the placebo and BTA groups \((p = 0.44)\).

In a sub-analysis, there was a statistically significant relationship between baseline CSS-GFL scores and percent change in patient-rated scores \((p = 0.03)\) and percent change in expert-rated scores \((p = 0.03)\), with a higher baseline CSS-GFL scores being associated with less improvement in mood. However, there was no significant association between baseline CSS-GFL scores and patient-rated response rates \((p = 0.13)\) or patient-rated remission rates \((p = 0.14)\). A similar insignificant relationship was found with expert-rated response rates \((p = 0.36)\) and remission rates \((p = 0.78)\).

**Discussion**

As expected and in concordance with the 3 individual trials, which all had positive findings, the pooled data showed with greater statistical power and a higher significance level, that a single treatment of BTA in the glabellar region can produce a strong reduction in the symptoms of major depression. With large effect sizes, this applied for all continuous and categorical efficacy outcome measures.

Our 3 studies are the only previously published randomized controlled trials on the use of BTA in the treatment of depression and they are well comparable with respect to the shared demographic and clinical baseline variables. The higher baseline scores in the BDI data in the study by EF and NER may indicate that patients with more severe depression were included in this study. However, when the data was pooled, there was no statistically significant difference in BDI scores or any other baseline variables between the placebo and BTA arms in this study.

In the pooled sample, patients who received BTA as a monotherapy and patients who received BTA as an adjunctive treatment improved equally. As such, there is potential for both treatment strategies. However, given the small sample size in each group \((n = 86; \text{monotherapy}, n = 49)\), and the results from EF and NER’s original study showing a superior remission rate for BTA augmentation vs. monotherapy \(36\% \text{ vs. } 21\%, \text{respectively}\), more studies are warranted before definitively concluding that monotherapy is as effective as augmentation.

There was no statistical difference in adverse events between the placebo and BTA groups and no severe reactions were reported \((p = 0.44)\). This excellent safety profile is in line with previous studies where BTA was injected in the glabella for aesthetic reasons \([23, 24]\). The safety and tolerability of BTA is one that has been studied extensively; however future studies on BTA for depression should continue to monitor safety data, as this is the largest study to date to report on safety for this specific indication.

The relationship between CSS-GFL scores and improvement in mood was inconclusive \(\text{statistically significant relationship to percent change in depression scores, but insignificant relationship to response and remission rates}\). Thus, there was a trend of higher baseline CSS-GFL scores being associated with less improvement in mood. This suggests that the presence of frown lines are not a good predictor of response, and that those with more severe frown lines will not necessarily have greater mood improvement than those with less severe frown lines. Perhaps more severe frown lines are an indicator of being depressed for a longer period of time, consequently making it a sign of “treatment resistance,” which is less likely to respond to any intervention. Future studies are warranted to investigate the correlation between frown scores and improvement in mood, and whether or not this could be used as a predictor of response.

We developed this new treatment approach based on the facial feedback hypothesis described in the introduction. This does not mean that proprioceptive facial feedback is the only conceivable mechanism of action. As discussed in the articles about the individual studies, it is possible that cosmetic changes and social feedback mechanisms related to the muscle-relaxing effect may be involved in mood elevation after BTA treatment. Moreover it is theoretically possible that direct effects on sensory neurons or even transport to and activity in the CNS may have a role \([25, 26]\). BTA as a treatment for depression offers some favorable properties. It is unique as a “one-time dose” intervention with an effect that lasts for months, it has an excellent safety and tolerability record, and it is already approved as a treatment of glabellar frown lines \([23, 24, 27]\). Furthermore, when taking into account the cost of brand medications and psychotherapy, it can be a cost-effective intervention for certain patient populations \([28]\).

With 3 positive randomized controlled trials and this pooled analysis of the individual patient data from all 3 of these trials, which broadens the validity of the results of the single trials, there is strong preliminary evidence for the efficacy of this treatment intervention and solid support for larger trials. The most important limitation of our findings, already discussed extensively in our previous papers, is the method-imminent difficulty to effectively blind participants to group allocation, associated with the obvious cosmetic effects of the treatment in most cases. However, unblinding due to side effects is a common problem in many bona fide double-blind trials \([29]\). Along with the comparably low placebo response in our pooled analysis, which may in part be related to unblinding, this may have inflated the differences between the BTA and the placebo group and complicates the estimation of the true effect size \([30]\). The low rate of improvement, response or remission in the placebo group may be related to this problem, as nocebo effects associated with the disappointment of being in the control group may counteract improvement in depression. The poor improvement in the placebo arm may also be explained by the high proportion of patients with chronic and partly treatment-resistant depression in the study samples. In these patients, the probability that
the spontaneous course leads to marked improvement or remission is low. Moreover, it is known that placebo responses are low in such patients [31]. In addition, our sample size was relatively small (n = 134). Although this generated statistically significant data with a large effect size, a larger sample size would have yielded more definitive results.

Strikingly few men were included in all 3 original studies. Also in the pooled sample the number of male participants was too low (n = 14) to determine the efficacy of the treatment in men. Consequently, there is a need for a study specifically investigating antidepressant effects of BTA in men.

Further studies are warranted to determine how BTA may be integrated into treatment algorithms for depression, be it as a sole or adjunctive treatment, in relapse prevention, in treatment-resistance or in antidepressant intolerance. In this context, psychomotor endophenotypes of depression (e.g., agitated depression) may also be predictors of response and allow for a stratified or personalized application of BTA [32]. Future research should also address if BTA can be used in the treatment of bipolar depression or other affective disorders. Moreover, other facial muscles like the depressor anguli oris (corners of mouth) or the mentalis (chin), which are also involved in depressed facial expression, may be treated in the future. A major challenge for future trials may be to find an adequate control condition. This may be an active comparator or even no treatment at all, which would at least circumvent nocebo effects that may be associated with undergoing an invasive procedure only to discover that one has received sham injections.

Should future studies replicate the safety and efficacy data from this pooled analysis, glabellar BTA injection may emerge as a novel therapeutic option for the treatment of depression.

Sources of Support

All 3 studies as well as the present pooled analysis were funded by private foundations or institutions, i.e., the Gottfried & Julia Bangerter-Rhyner-Stiftung, Bern, Switzerland, the Brain & Behavior Research Foundation, New York, USA, and the Chevy Chase Cosmetic Center, Chevy Chase, USA. None of these institutions were affiliated with pharmaceutical agencies.

Conflict of Interest

MAW received honoraria for talks from Merz, Eli Lilly and Novartis. THCK received honoraria for talks from Servier, Lundbeck, Eli Lilly and Dr. Schwabe. THCK received a grant from Trommsdorf. These activities were all unrelated to the study. MAW received a grant from the Asklepios Hamburg GmbH Forschungsförderung for related research. In April 2012, i.e., after conclusion and publication of the original study, MAW and THCK became members of the advisory board of Allergan. This activity ended in April 2014 for MAW and in September 2014 for THCK. In November 2012, i.e., after conclusion and as a result of the original study, MM became a consultant with Allergan, JSR’s spouse (MM) became a consultant for Allergan in 2012. EF has received a patent to treat depression with BTA. HR, BHK, SJ and NER declare no conflict of interest.

Affiliations

1 Department of Psychiatry, University of Texas at Austin, Dell Medical School, Austin, TX, USA (MM)
2 Department of Psychiatry, George Washington School of Medicine, USA
3 Department of Psychiatry, Social Psychiatry and Psychotherapy, Medical School Hannover, Germany (THCK, SJ)
4 Department of Analytics, Seton Family of Hospitals, Austin, TX, USA (HR)
5 Department of Dermatology, University of Texas at Austin, Dell Medical School, Austin, TX, USA (BH, JSR)
6 Capital Clinical Research Associates, Rockville, MD, USA (NER)
7 Georgetown Medical School, Washington, DC, USA (NER)
8 Asklepios Clinic North – Ochsenzell, Asklepios Campus Hamburg, Medical Faculty, Semmelweis University, Germany (SJ, MAW)
9 Psychiatric Clinics of the University of Basel, Switzerland (MAW)

References

1 Marcus M, Yasamy MT, van Ommeren M et al. Depression: a global public health concern. World Health Organization (WHO); fact sheet on depression 2012
17 Finzi E. The face of emotion: How botox affects our moods and relationships. Macmillan; 2013
20 Zimmerman M, Posternak MA, Chelminski I. Derivation of a definition of remission on the Montgomery-Asberg depression rating scale corresponding to the definition of remission on the Hamilton rating scale for depression. J Psychiatr Res 2004; 38: 577–582
27 Young SN. Single treatments that have lasting effects: some thoughts on the antidepressant effects of ketamine and botulinum toxin and the anxiolytic effect of psilocybin. J Psychiatry Neurosci 2013; 38: 78–83
30 Walsh BT, Seidman SN, Sysko R et al. Placebo response in studies of major depression: variable, substantial, and growing. JAMA 2002; 287: 1840–1847
32 Wollmer MA, Kalak N, Jung S et al. Agitation predicts response of depression to botulinum toxin treatment in a randomized controlled trial. Front Psychiatry 2014; 5: 36