Review Article

Deep Brain Stimulation Is Beneficial for Treatment-Resistant Depression: A Systematic Review and Meta-Analysis of Clinical Trials and Long-Term Follow-Ups

Ali Sepas[1,2,3], Hussein Alkafaji[4], Vedat Atici[4], Carsten R. Bjarkam[5], Torben Moos[4]

1. Department of Materials and Production, Aalborg University, Denmark; 2. Department of Quantitative Clinical Pharmacology, Research and Development, Novo Nordisk (United States), Seattle, United States; 3. Department of Mathematical Sciences, Aalborg University, Denmark; 4. Department of Health Science and Technology, Aalborg University, Denmark; 5. Department of Neurosurgery, Department of Clinical Medicine, Aalborg University Hospital, Aalborg, Denmark

Background: Most conventional treatments for MDD (major depressive disorder) fail to demonstrate significant results, which highlights an urgent need for novel treatments. DBS (deep brain stimulation) has shown great promise in several clinical trials. However, the efficacy of DBS, compared to conventional treatments for patients with TRD (treatment-resistant depression), and the targets with highest efficiency for treatment of TRD remain understudied objectives.

Methods: A systematic search in PubMed, ProQuest Dissertations and Theses Global, Embase, PsycINFO, and Scopus accompanied by a hand search in journals and grey literature was conducted in (January 10th, 2025). Any controlled trial with at least one group for comparison such as sham group was included. A random-effects meta-analysis was performed, quantitatively probing the efficacy of DBS vs sham/placebo for patients with TRD. A meta-analysis of response rate after long-term follow-up was also conducted. In addition, a meta-regression analysis was implemented to detect potentially moderating variables. Quality assessment was performed by utilizing the NOS (Newcastle Ottawa Scale) and Jadad scale.

Results: 1744 records were identified and screened for relevance, among which 14 were included in the quantitative analysis. Meta-analysis of the included studies revealed a medium-large effect size (SMD (standardized mean difference) Hedge's g = -0.51) in favor of active stimulation compared to sham. In the sensitivity analysis the effect size was statistically insignificant only when randomized controlled trials were included. The effect size and pooled response rate after long-term follow-up was = -1.12 and 56.14%, respectively.

Conclusion: The results indicate that DBS is potentially favorable for TRD compared to sham, as DBS exhibited greater efficacy compared to sham in both clinical trial and long-term follow-up phases. Further large-scale evidence-based studies, however, are required to substantially support these findings.

Corresponding authors: Ali Sepas, asepas20@student.aau.dk; Torben Moos, tmoos@hst.aau.dk

Introduction

MDD (major depressive disorder) is among the world's leading non-fatal causes of disability, affecting hundreds of millions worldwide[1]. Additionally, depressive patients may experience associated issues such as eating disorders, poor sleep, suicidal ideation, as well as problems with work, education, and social relations[2][3][4]. A large Canadian study[5] included 568,242 patients and found the treatment costs of depressed individuals to be more than three times that of comparably ill, non-depressed patients. The significantly increased costs can be explained by non–adherence to treatment and, possibly, by a greater number of doctor's visits by individuals in the depressed cohort[5][6], which illuminates a need for better handling of patients with depression.

The established treatment for MDD consists of psychotherapy and pharmacotherapy with antidepressants[6][7]. The medical treatment includes SSRIs (selective serotonin reuptake inhibitors), TCAs (tricyclic antidepressants), MAOIs (monoamine oxidase inhibitors), as well as various augmentation strategies, e.g., a TCA + Lithium[6]. While these strategies are valid for most clinically depressed patients, a large study (n = 3,671) found that only two-thirds of the patients with non-psychotic depression achieved remission following four consecutive stages of antidepressant treatment, each stage lasting approximately three months and consecutive stages being more aggressive than the preceding[8]. Additionally, patients who suffer from TRD (treatment-resistant depression) are usually referred to as ECT (electroconvulsive therapy)[6]. However, meta-analyses of published studies indicate that nearly half of the patients with a history of medication failure in the current depressive episode do not achieve clinical response following ECT[9][10]. Furthermore, the relapse of depressive symptoms following ECT remains a significant challenge[11][12], e.g. a recent study reported a relapse rate of 28.4%[13].

The failure of aggressive medical interventions for treating patients with severe TRD calls for novel approaches. One such approach is DBS (deep brain stimulation), which is a neurosurgical intervention that essentially functions as 'a pacemaker for the brain', delivering electrical impulses with the aim of

disrupting pathological neural activity[14]. In 2005, Mayberg and colleagues published the first clinical trial testing the effect of DBS on patients with TRD through stimulation of the SCC (subcallosal cingulate gyrus) of the brain (Brodmann area 25)[15].

These were followed by additional studies, which expanded the DBS targets for depression to include the MFB (medial forebrain bundle), the internal capsule (IC), the Acb (nucleus accumbens), the ITP (inferior thalamic peduncle) and the caudate nucleus (Cn),

among others[16][17]. The literature, however, remains undecided regarding the effects of DBS for treatment of TRD. Some studies report significant alleviation of depressive symptoms[18]. While others report no effect compared to sham[19], hence necessitating an up-to-date review of the literature.

This systematic review and meta-analysis reveal the first assessment of the efficacy of DBS for TRD. Our study also provides the first meta-analysis that reviews the findings of separate DBS treatment in RCTs (randomized controlled trials). Furthermore, we provide a meta-analysis of DBS treatment efficacy and the pooled response rate after long-term follow-up. As a secondary goal, our systematic review and meta-analysis also aim at detecting potential moderator variables that would influence the efficacy of DBS to lay the foundation for more specialized DBS treatment regimens for patients suffering from TRD.

Methods

Search Strategy

A systematic search was conducted (January 10th, 2025) in the following databases: PubMed, ProQuest Dissertations and Theses Global, Embase, PsycNet, and Scopus. An all-fields search was conducted in all databases except for ProQuest Dissertations and Theses Global and Scopus in where the search was conducted in all fields except for full text to limit the number of search outcomes to relevant hits only. An overview of the search terms can be seen in Table 1.

	AND			
OR	Psychiatric disorder	Disease stage	Intervention	Comparison
	*Depres** *Major depression* *Depressive disorder*	*Treatment resist** *Failed treatment* *Unresponsive* *Unmanageable* *Refractory treat** *Drug resist** *Intractable*	*DBS* *Deep brain stimulation*	*Sham* *Compar** *Placebo* *Set against* *Relation*

Table 1. A schematic illustration of the search string used for the systematic search. The full search string is presented in the supplementary materials.

Additionally, to maximize the number of potentially eligible articles, a hand search was performed in the following journals: Jama Network, Brain Stimulation, Depression & Anxiety, Biological Psychiatry, Neuropsychopharmacology, and Journal of Psychiatry & Neuroscience with the search term: "Treatment-resistant Depression and Deep Brain Stimulation". An additional search for ongoing trials in clinicaltrials.gov and a hand search in related articles such as citations from previously published systematic reviews and included studies were also conducted. The authors of the present manuscript also attempted to contact researchers identified in the literature search to obtain possible unpublished data, although no data was received. All identified articles were independently screened for inclusion by AS, HA and VA, and discrepancies were resolved internally with the senior authors CB and TM.

Selection Criteria

All potentially eligible articles were subjected to the following inclusion/exclusion criteria.

Inclusion:

- Articles include patients with TRUD (treatment-resistant unipolar depression as primary diagnosis), as defined by the authors of the included study

- Any form of a controlled trial with at least one comparison group such as sham, lithium treatment, another DBS target etc.
- Articles containing any form of DBS treatment for TRUD
- Only studies on humans, no age restriction
- Applied the proper instruments to measure depression, such as HAMD or any similar measurement tools
- Quantitative analysis of outcomes in the intervention and comparison group
- Articles from any country, articles in the English, Danish and Turkish languages were eligible
- Full text should be available for free online (including institutional access provided by the Aalborg University)

Exclusion:

- Articles include patients with anxiety, schizophrenia, and other psychiatric disorders except for unipolar depression
- Animal models of DBS for TRD
- Case studies, reviews, books, newspaper articles, posters, and letters to editors
- Exclusively qualitative analysis of outcomes
- Articles that do not include DBS treatment

Data Extraction

All the included articles assessed depression severity using various versions of the HAMD (Hamilton Depression Rating Scale) such as HAMD-17, HAMD-24, HAMD-29 or MADRS (Montgomery-Åsberg Depression Rating Scale). HAMD or MADRS scores were extracted. Only data from patients with TRUD was extracted when possible. Authors were contacted if data from bipolar patients could not be separated.

For each study, group means, and standard deviation (HAMD/MADRS score) were extracted, as well as publication year, national setting, sample size (number of patients included in the quantitative analysis), mean age of subjects at study entry, duration of active stimulation in the controlled trial phase, DBS target, and utilized depression scale. These were obtained from tables or calculated from individual data found in the published version of the studies included. For some studies, the tool WebPlotDigitizer was utilized to extract means and standard deviations from the graphs provided.

Quality Assessment

The methodological quality of each study was assessed using the JADAD scale for RCTs and the NOS (Newcastle-Ottawa scale) for non-RCTs (non-randomized controlled trials). Each scale returns a score, and based on this score, each article was given a ranking of either low, medium, or high quality. For RCTs, a maximum of five points was given in the following domains: randomization, blinding, and an account of outcomes in all patients. Studies were of high, moderate, or low quality if the score was ≥ 4, 2-3, or < 2, respectively.

The NOS assigns a maximum of nine points, assessing the categories of comparability, selection, and outcome. Studies were judged to be of high, medium, or low methodological quality if the score was > 5, 3-5, or < 3 respectively.

Statistical Analysis

The SMD (standardized mean difference), Hedge's g of each study was computed manually (accounting for paired and non-paired samples) based on the mean score and standard deviation of the depression scale in the intervention and comparison group post-treatment. Potential differences in sample sizes in the beginning and end of an intervention were accounted for through modified pooled standard deviation, where both sample sizes were included in addition to the sample size from the sham group, and the harmonic mean was used to account for changes in the intervention group. The effect sizes were synthesized by using the DerSimonian and Laird random effects model[20] in the statistics software R[21]. The random effects model was applied, since the studies were estimating related but different effect sizes due to variations in study design, surgical procedure, stimulation duration, depression rating scale, etc[22]. A two-way p-value of < 0.05 was considered significant.

Statistical heterogeneity was assessed by I^2, $I^2 < 50\%$ was considered low.

For sensitivity analysis, different meta-analyses were conducted where:

i. Only the results of RCTs for estimating the pooled effect size were included.
ii. Studies were excluded when the data from patients with unipolar and bipolar depression could not be separated.
iii. Studies were used when upfront randomization or blinded discontinuation were used.

iv. Only medium and high quality RCTs were included.

v. Only medium and high quality non-RCTs were included.

Meta-regression analysis was performed in R to identify potential moderator variables that might have influenced the effect sizes[21]. The included potential moderators were age, gender, stimulation length in the controlled trial phase, duration of the current depressive episode at the study entry, stimulation optimization duration, long-term follow-up duration, upfront randomization or blinded discontinuation. Effect sizes from two different DBS targets vs. sham in the studies by Raymaekers et al. (2017)[23] were considered separately in the meta-analysis.

Publication Bias

Funnel plots were used to evaluate bias and plot asymmetry. Additionally, Egger's test was utilized to quantitatively assess the asymmetry in the funnel plot. Finally, the non-parametric trim-and-fill method[24] was applied to detect whether any studies should be imputed in calculating the pooled effect sizes, possibly adjusting these effects.

Results

The initial search returned 1864 results (1082 after removal of duplicates); these were all screened for relevance based on title and abstract. Following this selection, 245 articles remained. Articles were then filtered by full-text screening based on the previously described inclusion/exclusion criteria. A total of 11 articles remained, which were all quantitatively analyzed. A PRISMA flow diagram illustrating the search procedure can be seen in Fig. 1.

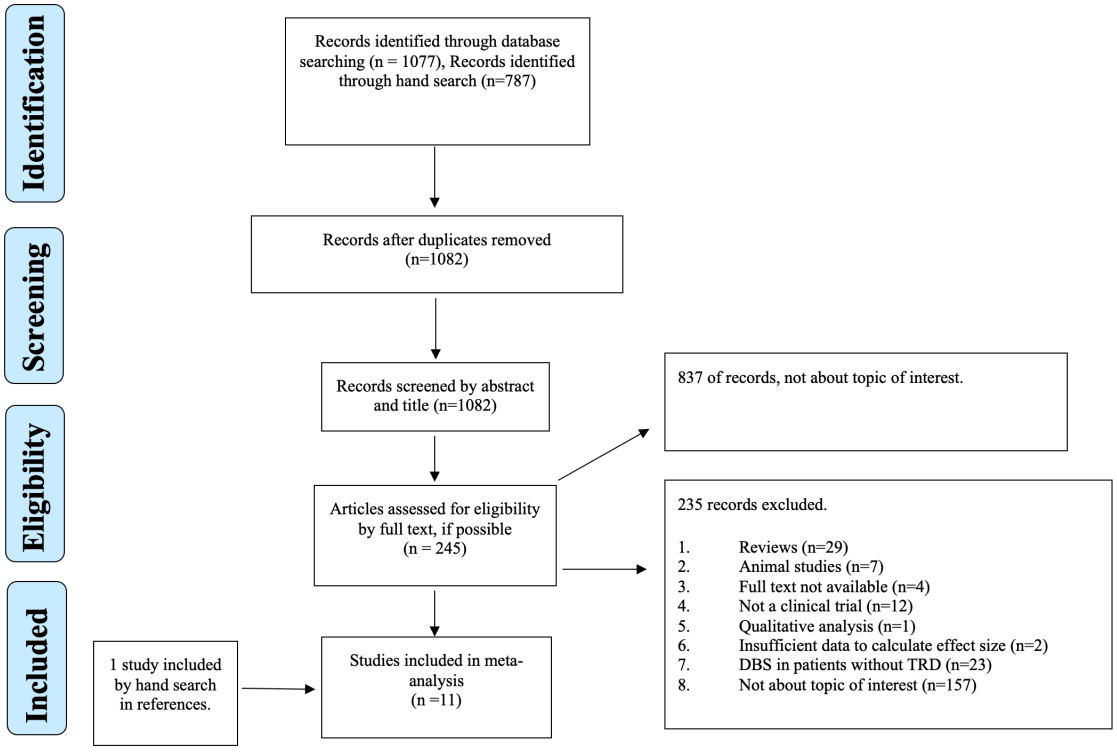

Figure 1. Flowchart

189 unique patients were included in the meta-analysis; the number of patients with unipolar depression was 187, and the number of patients with bipolar depression was 2. The average stimulation duration (trial phase) was 20.13 weeks. Some studies reported a follow-up extension period, where changes in pharmacotherapy were allowed concurrently with DBS or DBS parameters were adjusted to induce greater clinical efficacy.

The mean age of subjects across all studies was 49.4 ± 9.9 years. All studies were blinded, and, among the included studies, 4/11 utilized a crossover design with patients functioning as their own controls. Eight of the studies included were RCTs (Table 2). Inclusion criteria and the definition of TRD differed from study

to study. HAMD measures for inclusion ranged from ≥17 to ≥ 21 and MADRS from ≥ 21 to ≥ 26. Most studies defined TRD as a failure of four adequate antidepressant trials and ECT, analogous to stage V resistance described by Thase et al. in their seminal study of TRD[7]. All studies utilized a definition of TRD where at least two adequate antidepressant trials had failed to alleviate depressive symptoms· The definition of TRD in each study can be found in Supplementary Table S1.

No.	Author (Year)	National setting	Study design	Sample size (female %)	Mean/SD age at disease onset (years)	Mean age at study entry (years)	Stimulation target(s)	Stimulation duration (Trial phase) (weeks)	Optimization phase (weeks)	Follow-up duration after trial phase (weeks)	Depression scale	Hedge's g (g<0 = favors active stimulation) (g>0 = favors sham stimulation)	Quality Assessment
1.	Bergfeld et al. (2016)[18]	The Netherlands	RCT (crossover)	16 (NA)	37.8 ± (9.8)	53.1 ± 8.4	Ventral anterior limb of the IC (internal capsule)	6	52	-	HAMD-17	-1.55 [-2.35; -0.90]	High
2.	Coenen et al. (2019)[25]	Germany	RCT (parallel)	16 (38) [1 BPIID]	28.5 ± 9.8	51.6 ± 10.2	Superolateral MFB (medial forebrain bundle)	8	NA	44	MADRS	-0.65 [-1.59; 0.32]	High
3.	Dougherty et al., (2015) [19]	USA	RCT (parallel)	29 (45)	-	47.7 ± 12.0	VC/VS (Ventral Capsule/Ventral Striatum)	16	4	88	MADRS	0.11 [-0.60;0.82]	High
4.	Fenoy et al. (2018)[26]	USA	Controlled trial (single blinded)	6 (60)	15.2 ± 6.3	50.2 ± 10.2	Superolateral MFB	26	NA	52***	MADRS	-0.30 [-1.39; 0.80] **	Low
5.	Fenoy et al. (2022)[27]	USA	Controlled trial (single blinded)	4 (50)	22 ± 4.6	50.8 ± 3.5	Superolateral MFB	26	NA	600 (5 years) ***	MADRS	0.94 [-0.48; 2.28] **	Low
6.	Holtzheimer et al. (2012) [28]	USA	Controlled trial (single blinded)	10 (NA)	20.3 ± 5.6	40.0 ± 9.3	SCC	24	NA	144 (2 years) +	HAMD-17	-1.64 [-2.80;-0.83]	Medium
7.	Holtzheimer et al. (2017) [29]	USA	RCT (parallel)	85 (NA)	-	50.5 ± 9.7	SCC	26	10	144 (2 years)++	MADRS	-0.17 [-0.63; 0.29] +++	High
8.	Merkl et al. (2018)[30]	Germany	RCT (parallel)	8 (13) [1 BPID]	28.6 ± 9.7	48.3 ± 12.1	SCC	8	NA	144 (24 months) ******	HAMD-24	0.53 [-0.73;1.75]	High
9	Puigdemont et al. (2015) [31]	Spain	RCT (crossover)	5 (N/A)	23.2 ± 2.1	47.2 ± 13.0	SCC	13	NA	NA	HAMD-17	-0.53 [-1.76; 0.49]	High
10.	Ramasubbu et al. (2013) [32]	Canada	RCT (crossover)	4 (75)	50.2 ± 4.2	50.3 ± 3.6	SCC	12	12	26	HAMD-17	-0.49 [-1.88; 0.63]	High

No.	Author (Year)	National setting	Study design	Sample size (female %)	Mean/SD age at disease onset (years)	Mean age at study entry (years)	Stimulation target(s)	Stimulation duration (Trial phase) (weeks)	Optimization phase (weeks)	Follow-up duration after trial phase (weeks)	Depression scale	Hedge's g (g<0 = favors active stimulation) (g>0 = favors sham stimulation)	Quality Assessment
11.	Raymaekers et al.(2017) [23]	The Netherlands & Belgium	RCT (crossover)	Crossover 1 n= 6 Crossover 2 n=5 (NA)	35.3 ± 8.8	50.0 ± 5.2 (at cross over 1)	Ventral anterior limb of the IC & ITP (Inferior Thalamic Peduncle)	18	22 (5 months)	156 (3 years)	HAMD-17	IC vs sham: -1,24 [-2.11; -0.34] ITP vs sham: -0,58 [-1.81; 0.42] IC vs ITP (g < 0 = favors IC DBS (g > 0 = favors IT DBS): -0,21 [-1.42; 0.92]	High

Table 2. Summary of study characteristics

*BPID = Bipolar I Disorder, BPIID = Bipolar II Disorder, TRD = treatment-resistant depression, NA = not available, HAMD = Hamilton Depression Rating Scale, RCT: randomized controlled trial, MADRS: Montgomery-Åsberg Depression Rating Scale, SCC: Subcallosal cingulate gyrus, IC: internal capsule, MFB: medial forebrain bundle, * Data from the first year of the study was included, due to high dropout rate, **active outcome based on data from 26 weeks of stimulation. Although stimulation was continued for 52 weeks, patients were required to maintain the same medication for 6 months post-surgery,*** data from 1 year of follow-up included in subgroup meta analysis. ****active outcome = average of all active settings.***** Data from only 24 months was used for long-term follow-up analysis since data from all patients after 36 weeks (about 8 and a half months) were not available, ****** Data from only 24 weeks used for long-term follow-up analysis, data excluded from 28 weeks due to 50% drop out. + only 8 MDD patients remained after 2 years, not possible to separate BP from MDD patients for response rate analysis. ++ data from 2 years not included due to significant dropout rate in follow-up meta-analysis. For follow-up meta-analysis, data from n=82 was included, Data from 26 weeks (n= 90) at baseline were included in a primary meta-analysis of controlled results compared to long-term active (n= 30) baseline sham.*

The meta-analysis of 11 included studies (189 patients) comparing DBS to sham treatment revealed that active DBS led to a greater reduction in depressive symptoms compared to sham, with Hedges' g = −0.51, 95% CI [−0.97; −0.05], I^2 = 64%. The non-parametric trim-and-fill analysis did not adjust the effect size (Table 3, Fig. 2., Fig. 4(a)).

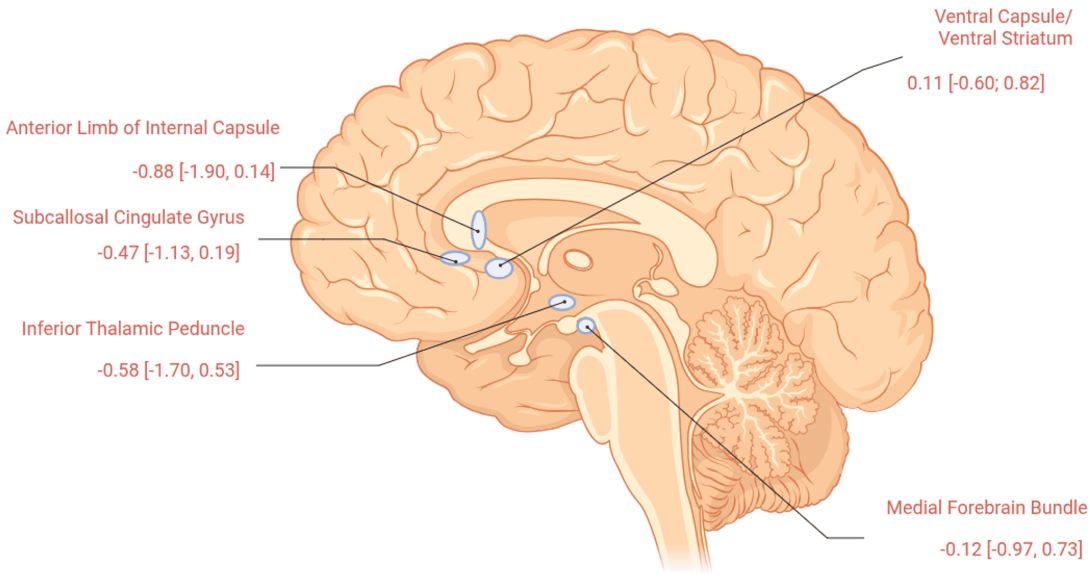

Figure 2. Stimulation Targets and Meta-analytic Results: Trial Phase

After excluding non-RCTs, the analysis of eight studies (169 patients) resulted in an effect size of g = −0.52, 95% CI [−1.01; −0.04] (Table 3), I^2 = 62%. Six studies utilized upfront randomization, and the combined pooled effect size was g = -0.35, 95% CI [-1.07, 0.36], I^2 = 59% (Table 3). In a separate comparison of internal capsule (IC) vs. inferior thalamic peduncle (ITP) DBS (5 patients), the effect size was g = −0.21, 95% CI [−1.38; 0.96] (Table 3). Egger's test indicated publication bias in the pooled analysis of all studies, that had a long-term follow-up phase (p = 0.03).

doi.org/10.32388/47TW5N

	Number of studies (subjects at baseline)	Hedge's g with 95% CI	p-value	Adjusted Hedge's g with 95% CI	Significant Moderators
DBS vs sham	11 (189)	-0.51 [-0.97; -0.05]	< 0.05	No adjustments	None
DBS vs sham (only RCTs)	8 (169)	-0.52 [-1.01, -0.04]	< 0.05	-0.19 [-0.76, 0.39]	None
DBS vs sham (only medium to high quality non-RCTs)	1 (10)	-1.64 [-2.80;-0.83]	< 0.05	No adjustments	None
DBS vs sham (excluding studies with bipolar patients)	9 (165)	-0.58 [-1.10; -0.06]	< 0.05	No adjustments	None
DBS vs sham upfront randomization	6 (49)	-0.35 [-1.07, 0.36]	> 0.05	-0.80 [-1.62, 0.015]	None
DBS vs sham blinded discontinuation	5 (140)	-0.65 [-1.30, 0.01]	> 0.05	-0.14 [-0.93, 0.66]	None
DBS SCC vs sham	5 (112)	-0.47 [-1.13, 0.19]	> 0.05	No adjustments	None
DBS SCC vs sham (only RCTs)	4 (102)	-0.18 [-0.56, 0.21]	> 0.05	No adjustments	None
DBS SCC vs sham (excluding studies bipolar patients)	4 (104)	-0.65 [-1.33, 0.04]	> 0.05	-0.24 [-0.99, 0.51]	None
DBS IC vs sham	3 (51)	-0.88 [-1.90, 0.14]	> 0.05	-0.88 [-1.904, 0.14]	None
DBS MFB vs sham	3 (26)	-0.12 [-0.97, 0.73]	> 0.05	-0.12 [-0.97, 0.73]	None
DBS MFB vs sham (RCT)	1 (16)	-0.65 [-1.60, 0.32]	> 0.05	No adjustments	None
DBS MFB vs sham (excluding bipolar patients)	2 (10)	0.25 [-0.96, 1.45]	> 0.05	No adjustments	None
ITP DBS vs. sham	1(5)	-0.58 [-1.70, 0.53]	> 0.05	No adjustments	None

Table 3. Summary of meta-analysis and sensitivity analysis (trial phase)

**Adjusted Hedge's g values were calculated in STATA following the application of non-parametric trim-and-fill analysis for reducing publication bias by imputing more studies.*

The overall pooled response rate was r = 0.56, 95% CI [0.36; 0.76], I^2 = 92% (Table 4, Fig. 5(a)). When non-RCTs were excluded, the response rate adjusted to r = 0.48, 95% CI [0.25; 0.70], I^2 = 91% (Table 4, Fig. 6). Furthermore, upfront randomization vs. blinded discontinuation as a moderator was also significant in DBS vs sham (excluding studies with bipolar patients, response rate) (Table 4). Eight studies (166 patients) conducted open-label long-term follow-up, yielding a pooled effect size of g = –1.12, 95% CI [–1.89; –0.36], I^2 = 74% (Table 5, Fig. 3, Fig. 6).

	Number of studies (subjects)	Hedge's g with 95% CI	p-value	Adjusted Hedge's g with 95% CI	Significant Moderators
DBS vs sham	8 (166)	0.56 [0.36; 0.76]	< 0.01	0.29 [0.0184, 0.55]	None
DBS vs sham (only RCTs)	6 (146)	0.48 [0.25, 0.70]	< 0.010.26 [-0.01, 0.53]		None
DBS vs sham (excluding studies with bipolar patients)	6 (142)	0.54 [0.30, 0.77]	< 0.05	0.28 [-0.02, 0.58]	Upfront randomization vs blinded discontinuation
DBS vs sham (only medium to high quality non-RCTs)	1 (10)	0.92 [0.65, 0.99]	< 0.01	No adjustments	None
DBS vs sham upfront randomization	4 (44)	0.81 [0.56, 1.06]	< 0.01	No adjustments	None
DBS vs sham blinded discontinuation	4 (122)	0.37 [0.20, 0.54]	< 0.01	0.31 [0.09, 0.52]	None
DBS SCC vs sham	4 (107)	0.50 [0.19, 0.82]	< 0.01	0.28 [0.018, 0.55]	None
DBS SCC vs sham (only RCTs)	3 (97)	0.28 [0.18, 0.38]	> 0.05	0.26 [-0.01, 0.53]	None
DBS SCC vs sham (excluding studies bipolar patients)	3 (99)	0.56 [0.15, 0.96]	< 0.01	0.26 [-0.19, 0.70]	None
DBS IC vs sham	2 (35)	0.46 [-0.01, 0.93]	p > 0.05	No adjustments	None
DBS MFB vs sham	2 (26)	0.79 [0.59, 1.00]	p < 0.05	No adjustments	None
DBS MFB vs sham (RCT)	1 (16)	0.88 [0.64, 0.97]	p < 0.05	No adjustments	None
DBS MFB vs sham (excluding bipolar patients)	1 (10)	0.66 [0.35, 0.88]	p < 0.05	No adjustments	None

Table 4. Summary of response rate meta-analysis

**Adjusted Hedge's g values were calculated in STATA following the application of non-parametric trim-and-fill analysis for reducing publication bias by imputing more studies.*

For the long-term follow-up, the exclusion of non-RCTs resulted in six studies, yielding g = −0.80, 95% CI [−1.29; −0.30], $p < 0.05$, $I^2 = 42\%$ (Table 5). The trim-and-fill analysis adjusted the effect size to g = −0.41, 95% CI [−0.98; 0.16]. When studies involving bipolar patients were excluded, the meta-analysis of DBS vs. sham produced an effect size of g = -1.08, 95% CI [−2.12; −0.04], $p < 0.05$, $I^2 = 68\%$ (Table 5). Five studies specifically assessed SCC DBS vs. sham, with an effect size of g = −0.47, 95% CI [−1.13; 0.19], $p > 0.05$, $I^2 = 58\%$ (Table 5, Fig. 3, Fig. 6(b)). Excluding non-RCTs in the analysis of SCC DBS vs. sham adjusted the result to g = −0.18, 95% CI [−0.56; 0.21], $I^2 = 47\%$ (Table 5). For medial forebrain bundle (MFB) DBS vs. sham (26 patients), the effect size was g = −0.12, 95% CI [−0.97; 0.73], $I^2 = 43\%$ (Table 5, Fig. 3, Fig. 6(d)). Since all the RCTs were of high quality, no further sensitivity analysis was required. Accordingly, only high quality RCTs were included in the relevant analyses. Age was found to be a significant moderator in the analysis of DBS vs. sham (all studies included, long-term follow-up), and DBS vs. sham (excluding studies with bipolar patients, long-term follow-up). Upfront randomization vs. blinded discontinuation was a significant moderator in the analysis: DBS vs. sham (only RCTs, long-term follow-up) (Table 5). The most relevant funnel plots are presented in (Figs. 7-9).

	Number of studies (subjects)	Hedge's g with 95% CI	p-value	Adjusted Hedge's g with 95% CI	Significant Moderators
DBS vs sham	8 (166)	-1.12 [-1.89; -0.36]	< 0.05	−0.59 [-1.64, 0.46]	Age
DBS vs sham (only RCTs)	6 (146)	-0.80 [-1.29, -0.30]	< 0.05	-0.41 [-0.98, 0.16]	Upfront Randomization or Blinded discontinuation
DBS vs sham (excluding studies with bipolar patients)	6 (142)	-1.08 [-2.12; -0.04]	< 0.05	No adjustments	Age
DBS vs sham (only medium to high quality non-RCTs)	1 (10)	-1.64 [-2.8; -0.83]	< 0.05	No adjustments	None
DBS vs sham upfront randomization	4 (44)	-1.73 [-3.29, -0.18]	< 0.05	No adjustments	None
DBS vs sham blinded discontinuation	4 (122)	-0.49 [-0.86, -0.13]	< 0.05	-0.40 [-0.75, -0.05]	None
DBS SCC vs sham	4 (107)	-1.42 [-3.07, 0.23]	> 0.05	-0.43 [-2.27, 1.40]	None
DBS SCC vs sham (only RCTs)	3 (97)	-0.47 [-0.92, -0.01]	< 0.05	-0.37 [-0.78, 0.0371]	None
DBS SCC vs sham (excluding studies bipolar patients)	3 (99)	-1.63 [-3.94, 0.69]	> 0.05	-1.63 [-3.95, 0.69]	None
DBS IC vs sham	2 (35)	-0.73 [-1.61, 0.16]	> 0.05	No adjustments	None
DBS MFB vs sham	2 (26)	-1.09 [-2.52, 0.33]	> 0.05	o adjustments	Nne
DBS MFB vs sham (RCT)	1 (16)	-1.80 [-2.77, -0.81]	< 0.05	No adjustments	None
DBS MFB vs sham (excluding bipolar patients)	1 (10)	-0.35 [-0.53, 1.45]	> 0.05	No adjustments	None

Table 5. Summary of meta-analysis and sensitivity analysis (long term follow-up phase)

**Adjusted Hedge's g values were calculated in STATA following the application of non-parametric trim-and-fill analysis for reducing publication bias by imputing more studies. Number of subjects in baseline is included.*

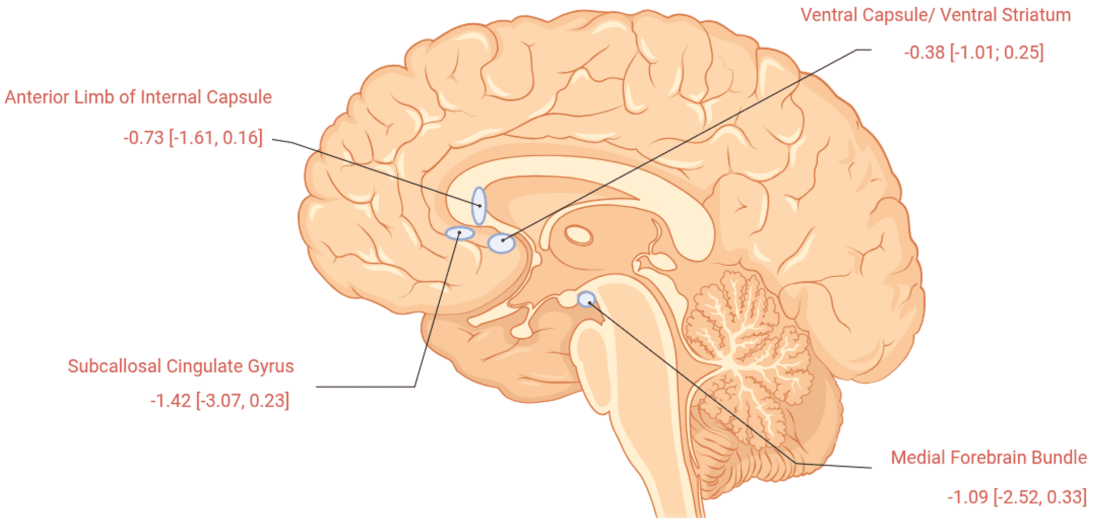

Figure 3. Stimulation Targets and Meta-analytic Results: Long-term Follow-up

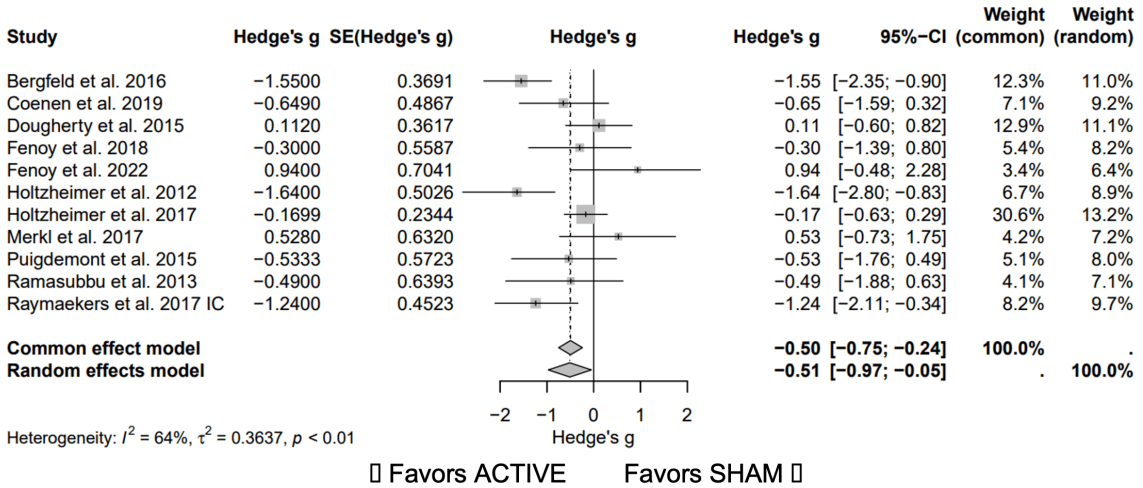

Figure 4(a). DBS vs. sham (all studies included)

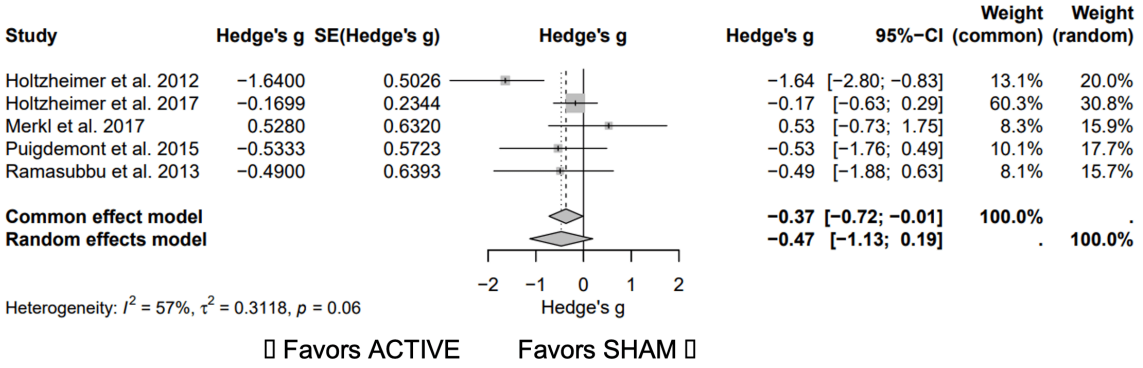

Figure 4(b). SCC DBS vs. sham (all studies included)

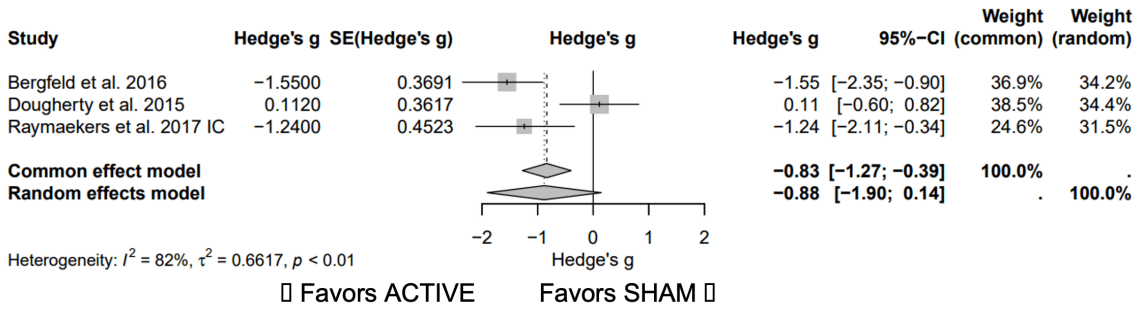

Figure 4(c). IC DBS vs. sham (all studies included)

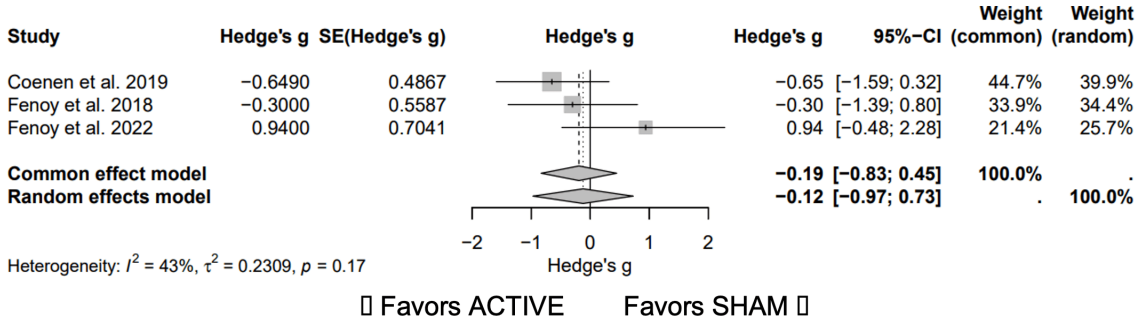

Figure 4(d).MFB DBS vs. sham (all studies included)

Figure 4. Forest plots comparing the effect of active DBS to sham (trial phase).

CI = confidence interval, Acb = Nucleus accumbens, Cn = Caudate nucleus, IC = internal capsule, ITP = inferior thalamic peduncle.

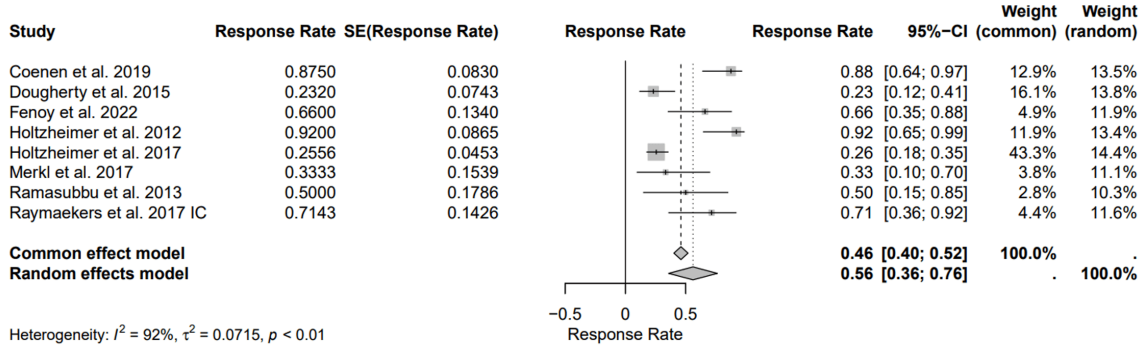

Figure 5(a). DBS vs. sham (all studies included)

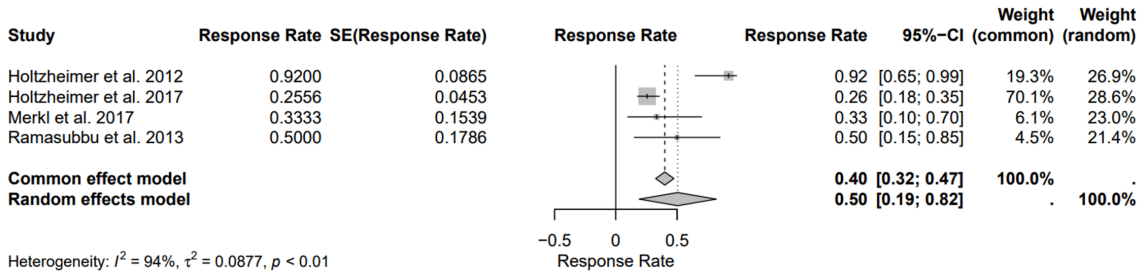

Figure 5(b). SCC DBS vs. sham (all studies included)

doi.org/10.32388/47TW5N

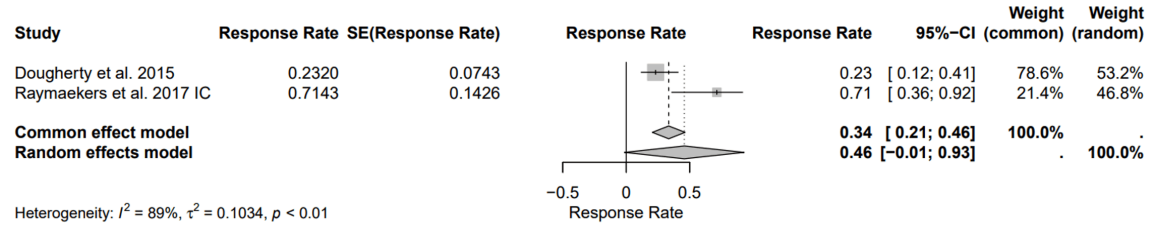

Figure 5(c). IC DBS vs. sham (all studies included)

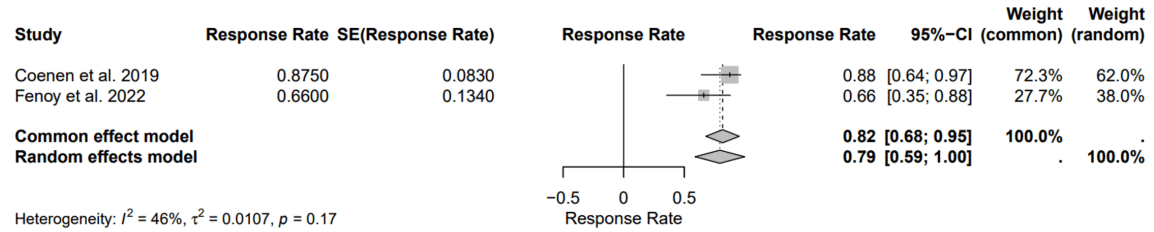

Figure 5(d). MFB DBS vs. sham (all studies included)

Figure 5. Forest plots of response rate

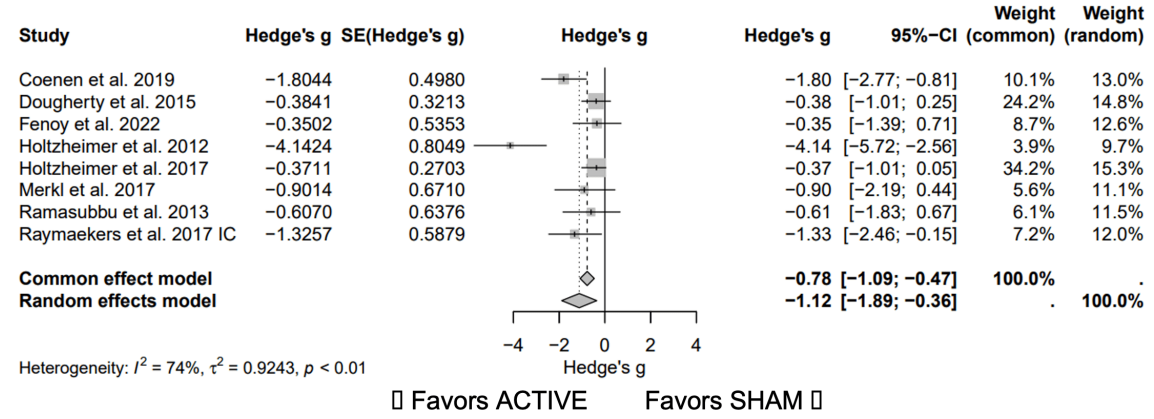

Figure 6(a). DBS vs. sham (all studies included)

doi.org/10.32388/47TW5N

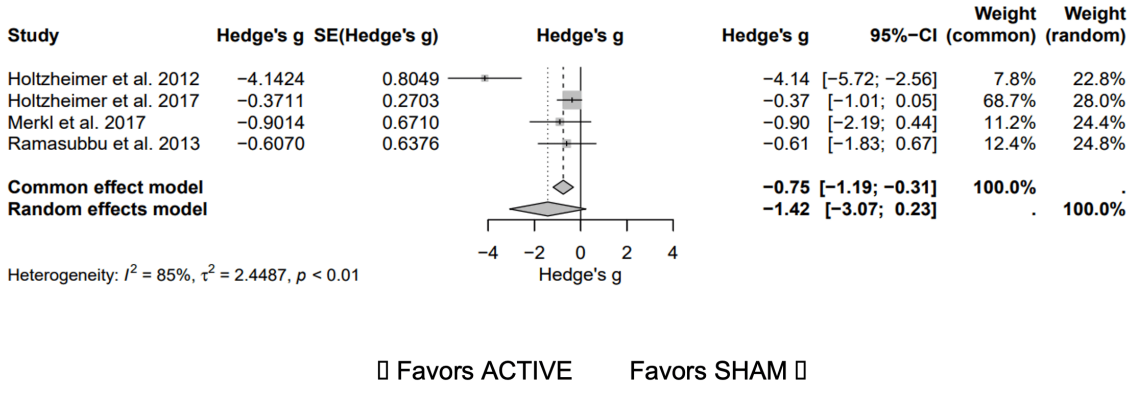

Figure 6(b). SCC DBS vs. sham (all studies included)

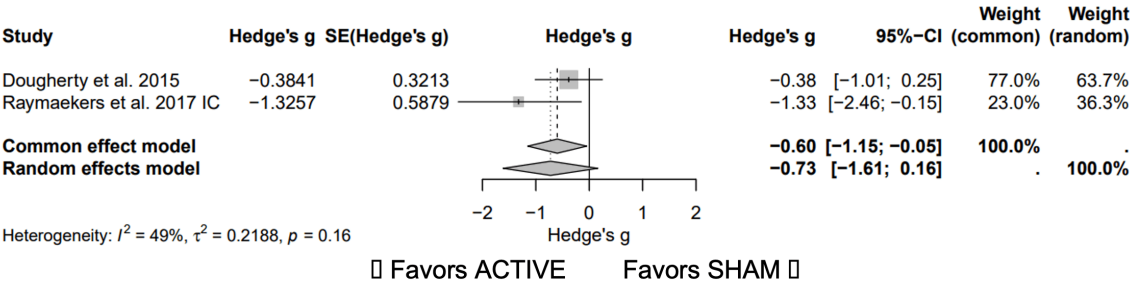

Figure 6(c). IC DBS vs. sham (all studies included)

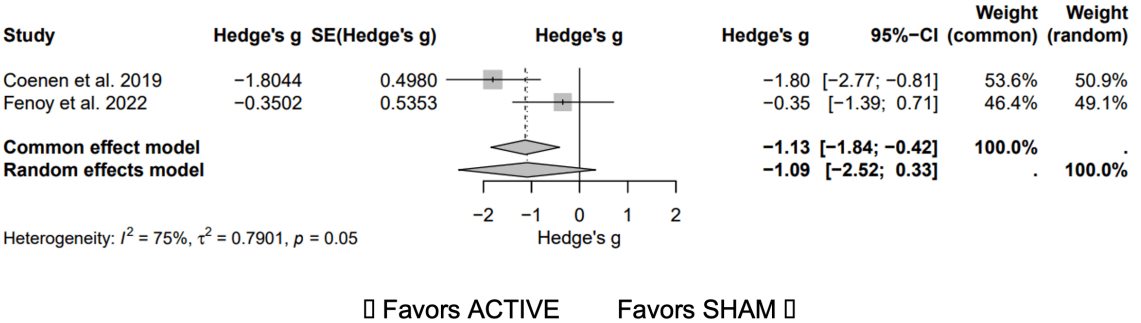

Figure 6(d). MFB DBS vs. sham (all studies included)

Figure 6. Forest plots comparing the effect of active DBS to sham (long term follow-up phase)

doi.org/10.32388/47TW5N

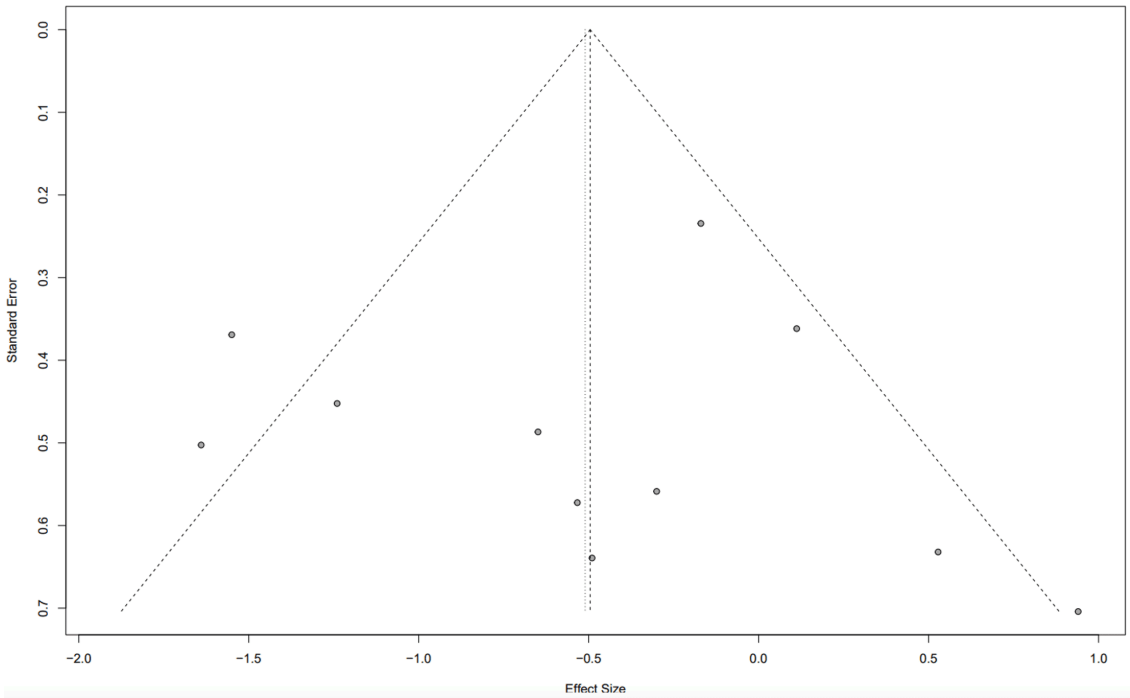

Figure 7(a). DBS vs. sham (all studies included)

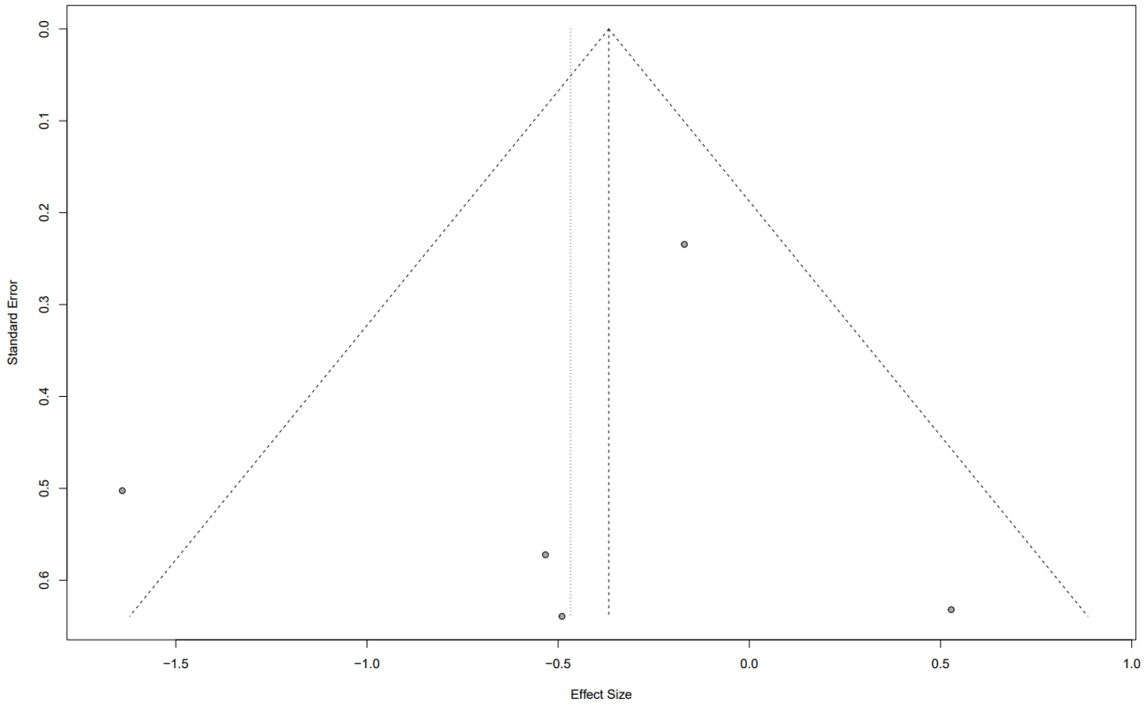

Figure 7(b). SCC DBS vs. sham (all studies included)

doi.org/10.32388/47TW5N

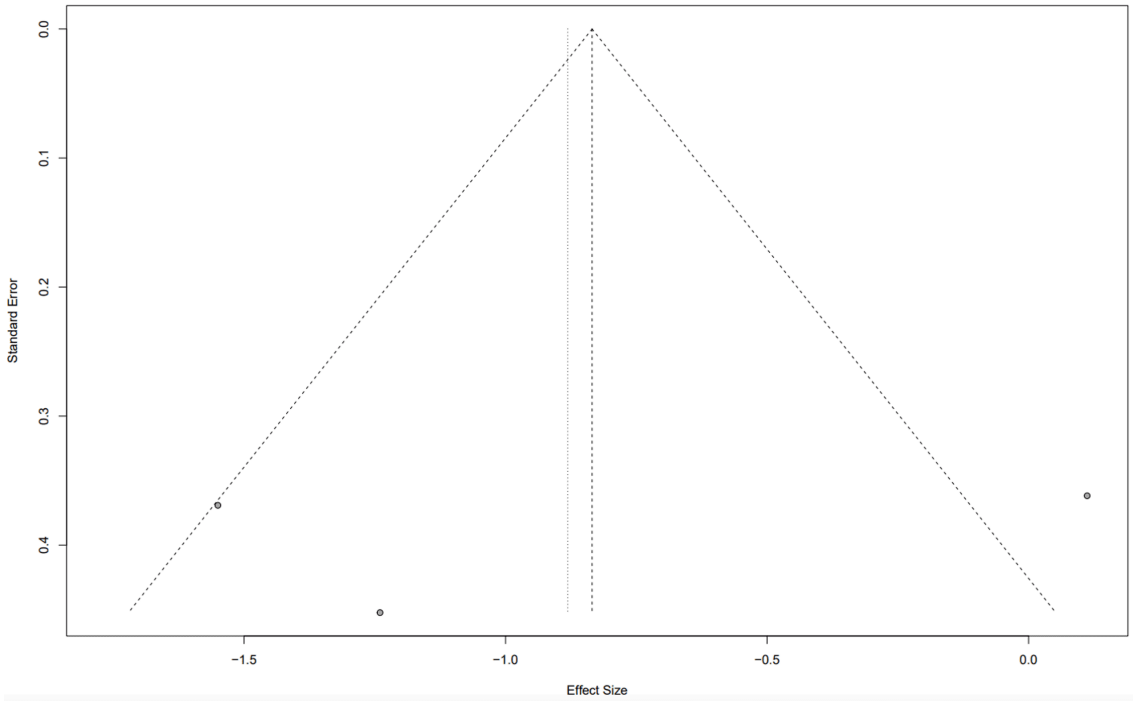

Figure 7(c). IC DBS vs. sham (all studies included)

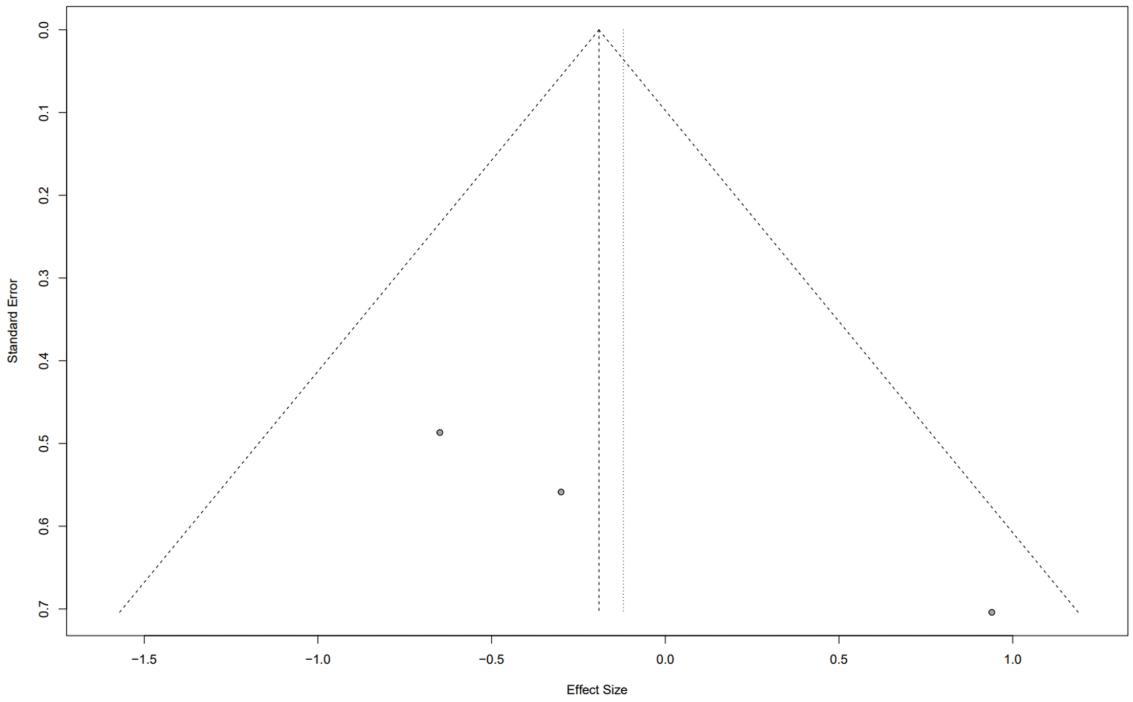

Figure 7(d). MFB DBS vs. sham (all studies included)

Figure 7. Funnel plots of DBS vs. sham (trial phase)

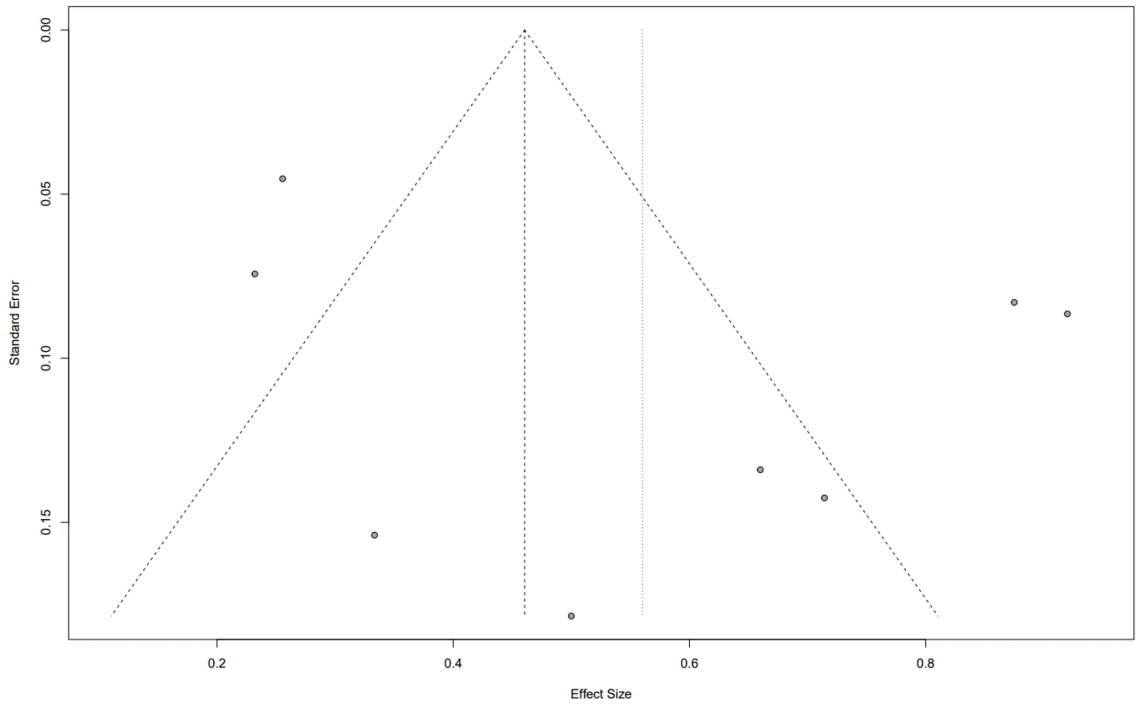

Figure 8(a). DBS vs. sham (all studies included)

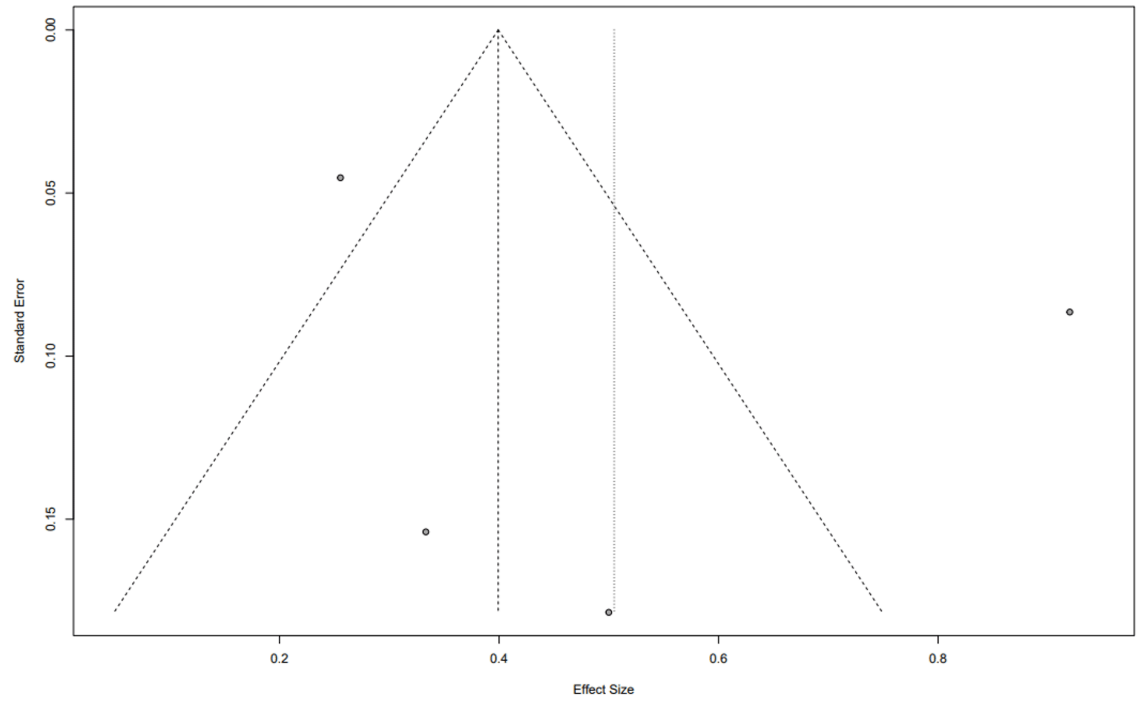

Figure 8(b). SCC DBS vs. sham (all studies included)

doi.org/10.32388/47TW5N

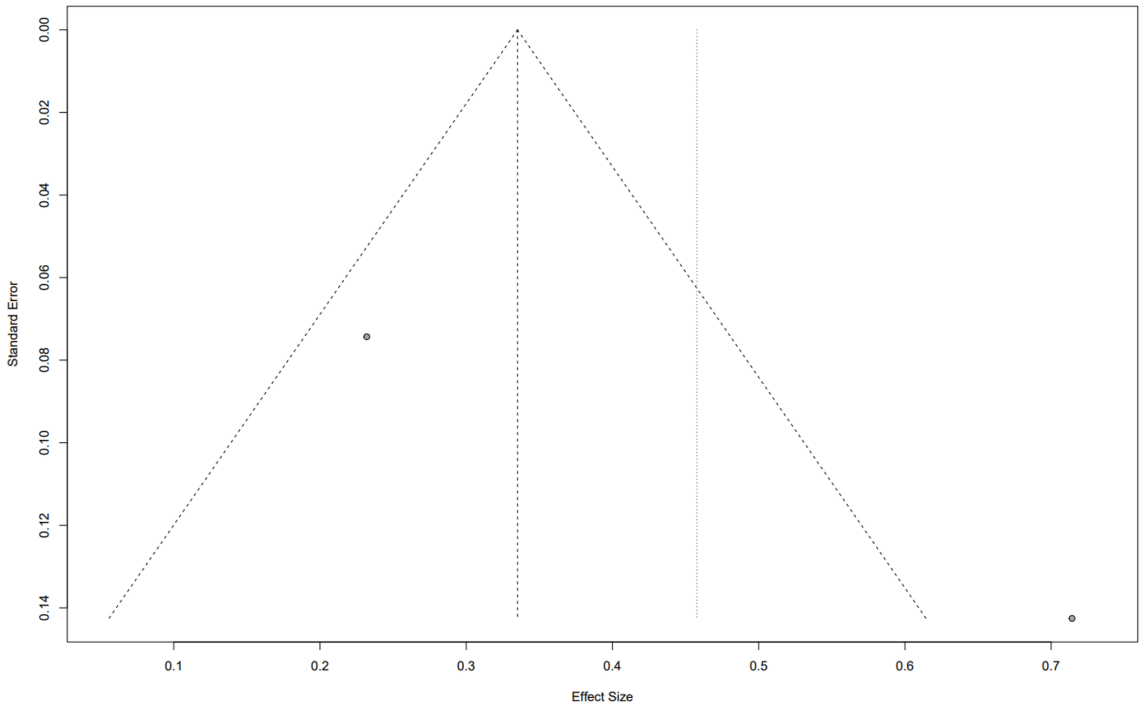

Figure 8(c). IC DBS vs. sham (all studies included)

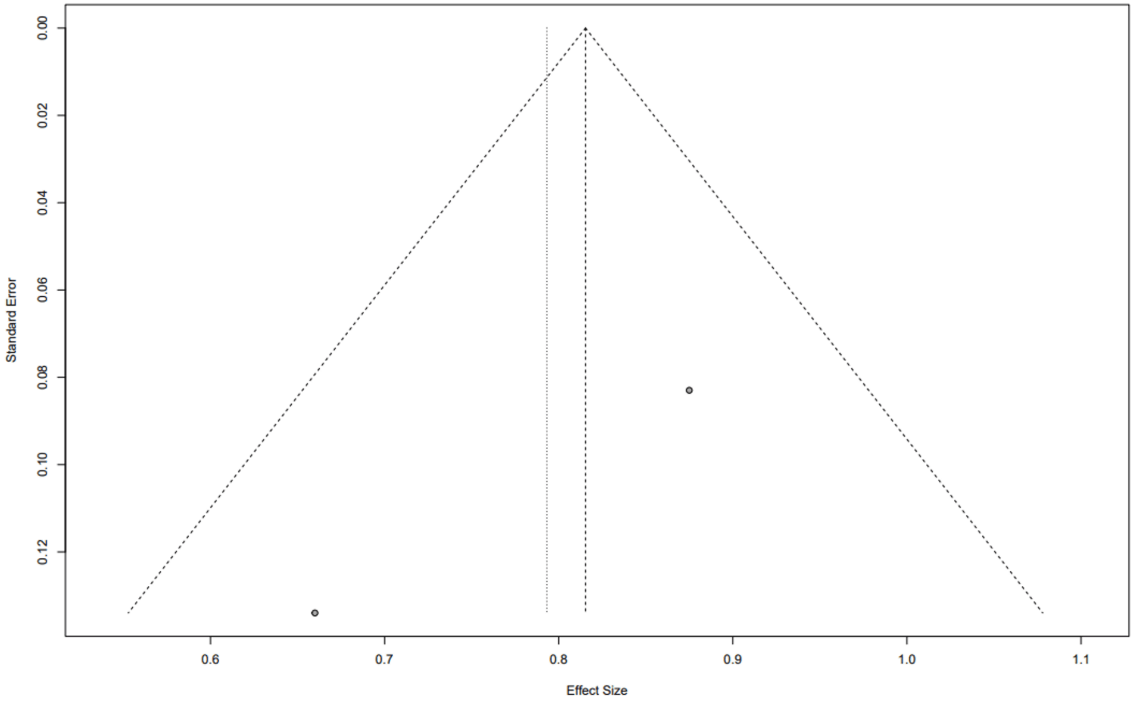

Figure 8(d). MFB DBS vs. sham (all studies included)

Figure 8. Funnel plots of response rate

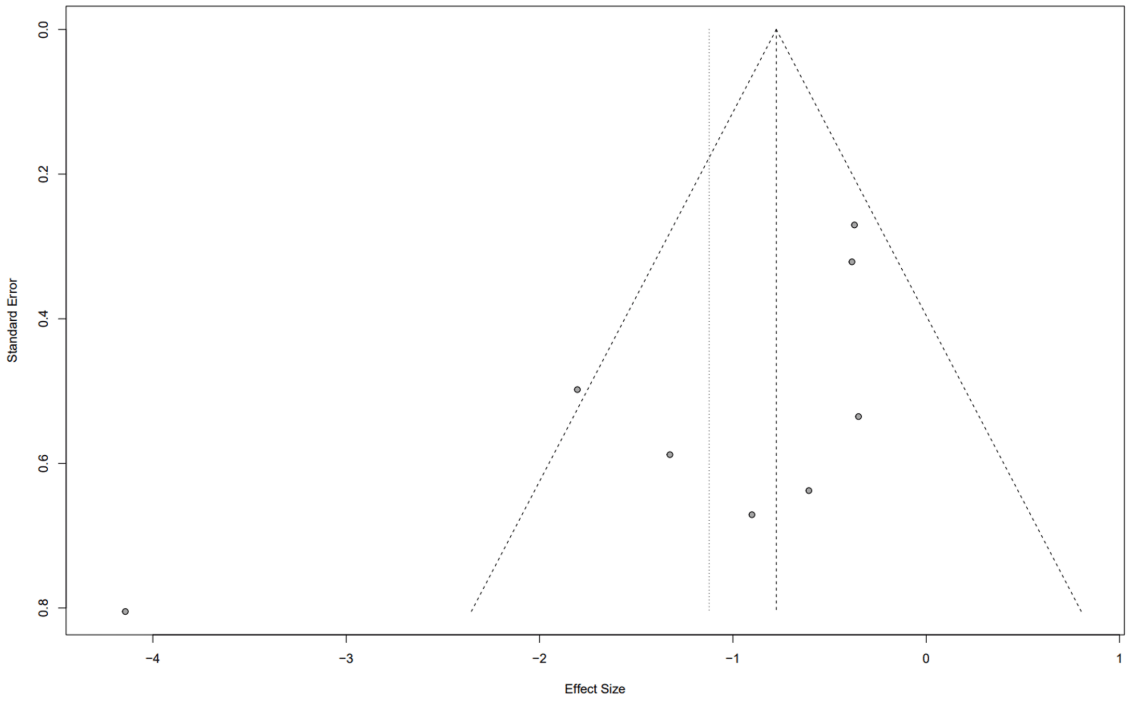

Figure 9(a). DBS vs. sham (all studies included)

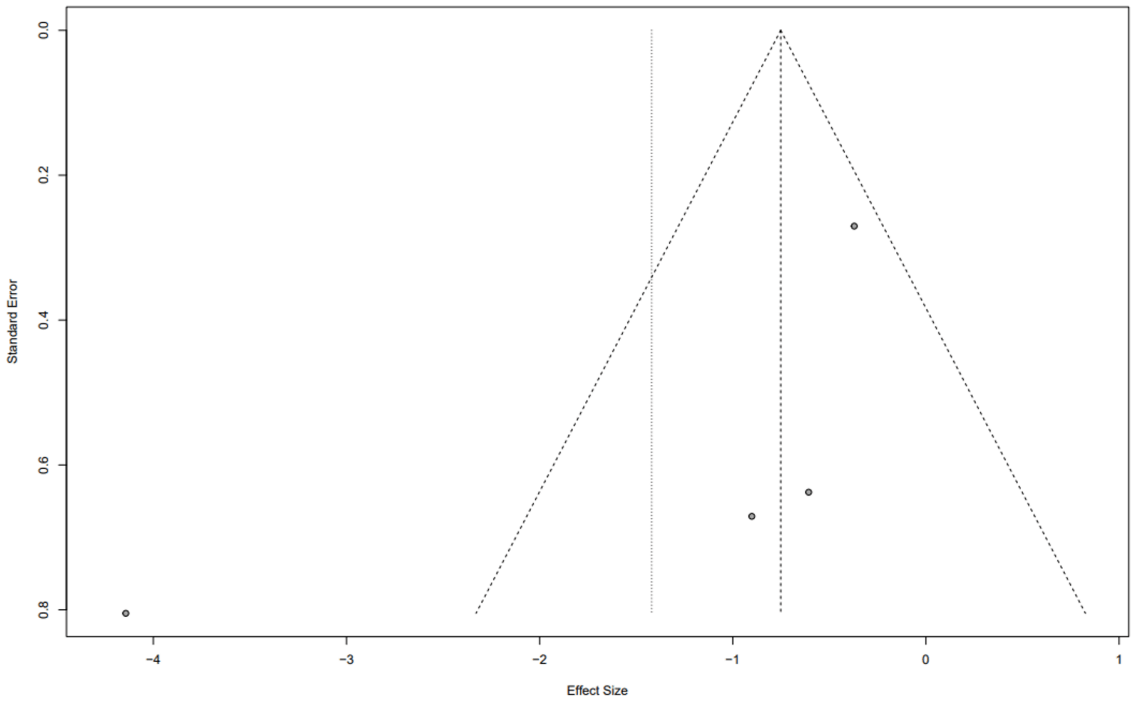

Figure 9(b). SCC DBS vs. sham (all studies included)

doi.org/10.32388/47TW5N

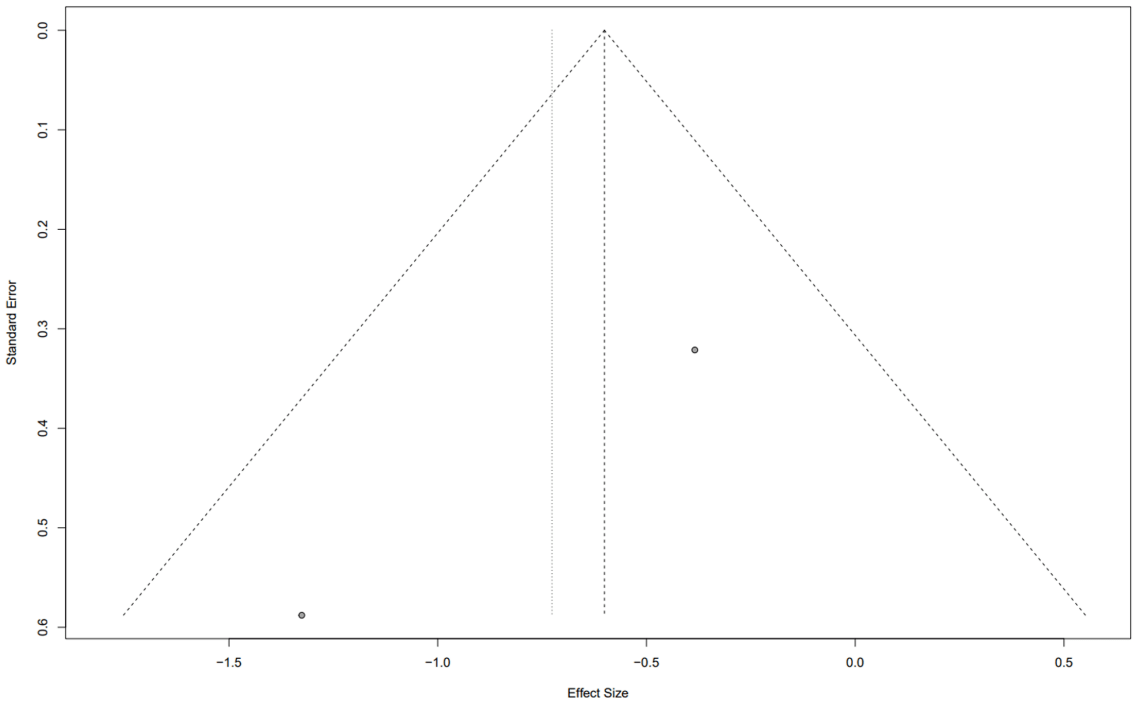

Figure 9(c). IC DBS vs. sham (all studies included)

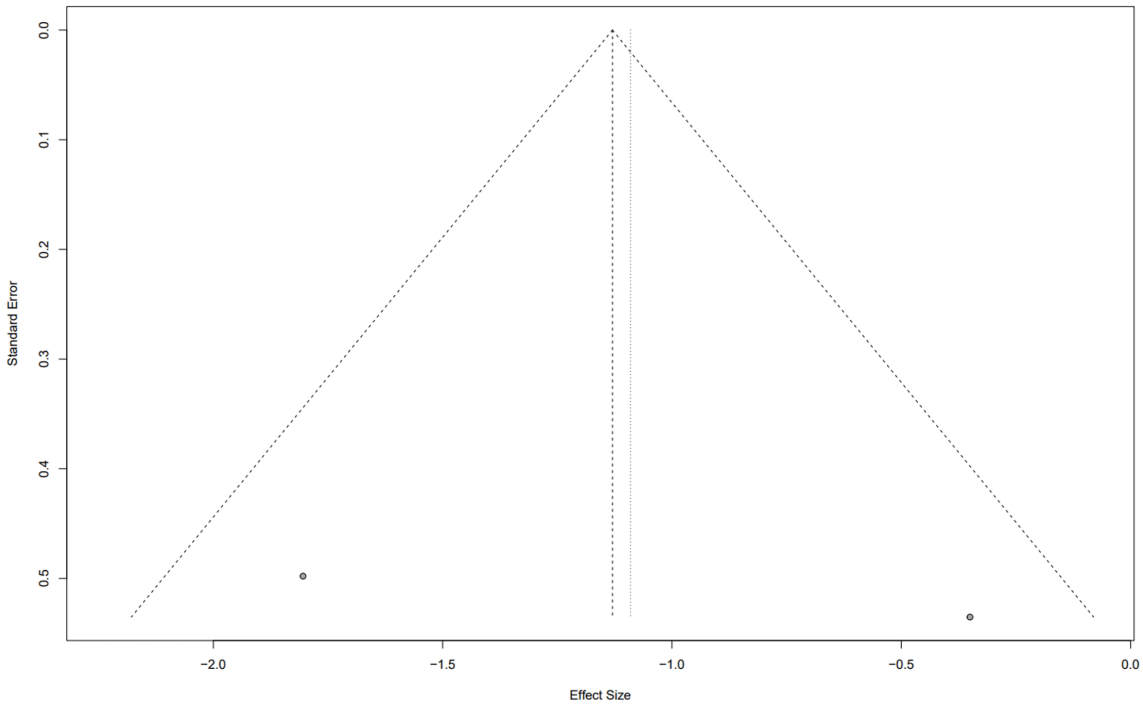

Figure 9(d). MFB DBS vs. sham (all studies included)

Figure 9. Funnel plots of active DBS vs. sham (long-term follow-up phase)

Discussion

This paper analyzed the efficacy of DBS as a treatment for TRD [16][33][34][35]. Corroborating earlier reports, DBS significantly alleviated depressive symptoms in TRD patients. Meta-analysis yielded a medium-large[36] effect size comparing DBS to sham, with a Hedges' g = −0.51, 95% CI [−0.97; −0.05]. Additionally, meta-analysis of DBS vs. sham (excluding non-RCTs) yielded a SMD of g = −0.52, 95% CI [−1.01; −0.04], which suggests that DBS represents a possible alternative when conventional treatment strategies fail. Analysis of the different targets showed that SCC as DBS target has greater efficacy than sham g = −0.47, 95% CI [−1.13; 0.19]. Furthermore, the sensitivity analysis of SCC DBS vs sham (excluding non-RCTs) did not yield significant difference g = −0.18, 95% CI [−0.56; 0.21]. Consequently, future RCTs with bigger sample sizes are required to assess whether SCC DBS outperforms sham surgery. Most of the RCTs had small sample sizes, resulting in wider confidence intervals and greater uncertainty in interpreting the results. Patient-specific stimulation design in RCTs, originally guided by MRI/CT scans[30] and now by diffusion tractography[27], vary in both target location and stimulation settings (amplitude, pulse width, frequency)[26][31]. The chosen settings critically influence clinical outcomes, highlighting the need for individualized treatment protocols.

Age was found to be a significant moderator in the analysis of DBS vs. sham (all studies included, long-term follow-up), and upfront randomization vs. blinded discontinuation as a moderator variable was also significant in DBS vs. sham (only RCTs, long-term follow-up). These findings suggest that age and stimulation strategy could play a role in how well patients respond to DBS over time. However, these findings are based on a few studies, and no moderator was consistently significant across all sensitivity analyses; consequently, current evidence remains too limited for firm conclusions.

Prior meta-analyses have also assessed the efficacy of DBS in TRD. All have demonstrated significantly better results with active stimulation compared to sham[16][33][34][35]. A meta-analysis of Hitti and colleagues[16] yielded similar results to the current study. The authors found DBS to be better than sham overall in the trial phase. The differences can be attributed to the inclusion of the study by Fenoy et al. and accounting for changes in sample sizes during the time course of some of the trials[27]. Furthermore, in studies where patients underwent sham stimulation before active stimulation, we used the difference between sham and corresponding active stimulation endpoints to account for possible insertional effects.

Additionally, when calculating the effect sizes for individual studies, no considerations regarding paired or unpaired samples were made. The differing efficacies reported by literature may be attributed to multiple factors[16]. Firstly, the neurophysiological properties of depression vary from individual to individual, making it difficult to pinpoint which target(s) is ideal for DBS[37]. Current studies do not fully account for these neurobiological differences. Secondly, it is yet to be explored whether non-responders to DBS of one brain target region may benefit from stimulation of an alternative target[37]. However, the neurosurgical nature of DBS treatment limits these possibilities. Thirdly, differing efficacy may relate to the accuracy of measurement of treatment outcomes. As noted by Rabin et al.[38], suboptimal measures of outcomes stand in the way of truly recognizing signs of effective treatment. While depression rating scales are generally acknowledged for assessing depression, they do present some inadequacies, especially when synthesizing data in meta-analyses[39].

For instance, multiple versions of the HAMD scale were used in the literature, despite all versions being tailored to different uses and including different domains of assessment. This may lead to overestimation of the importance of certain factors, depending on which scale is used. This, of course, extends to differences between HAMD and other rating scales, such as MADRS.

A great advantage of DBS is that active/sham stimulation can be compared within the same patient. This reduces inter-subject variability and allows for greater statistical power with fewer participants. This advantage was reflected in the included trials that utilized crossover design. However, the two largest of the included trials both used a parallel design analysis[19][29]. Both studies reported substantially lower efficacy of active stimulation compared to sham. A possible explanation for this discrepancy, as noted in the study by Raymaekers et al.[23], is that subjects were able to correctly "guess" whether stimulation was on or off, which suggests that adequate blinding was not achieved.

Limitations and direction for future research

Most studies utilized various versions of HAMD, which have been criticized for being flawed and incapable of depicting depression severity, as well as some of the main aspects of depression, e.g. feelings of worthlessness and anhedonia[40]. Furthermore, some evidence suggests that the total score of HAMD, which was utilized in the present study, does not adequately reflect depression severity[39]. However, a more recent critical review of various forms of HAMD validated the scale and reported it as being sensitive[39]. MADRS was utilized as well in some studies; however, this scale has been criticized for being less sensitive in detecting treatment effects compared to HAMD[41].

Another limitation may be the lack of a coherent definition for TRD in scientific literature. Some studies resorted to defining TRD as the failure of response to multiple adequate trials of various antidepressants, while others utilized a staging approach such as the Thase-Rush five-stage classification. Nevertheless, further evidence is required to assess the efficacy of such staging models[42]. A consensus that seems to emerge is to define TRD as a failure of response to at least two adequate trials of antidepressants[34][43].

Future studies should consider utilizing large-scale double-blinded RCT designs and utilize a better definition of TRD, where consensus is emerging. Additionally, researchers should consider the most appropriate depression scale based on their approach for assessing treatment efficacy; for instance, if the goal is to assess the different severity levels of depression, then HAMD-17 can be utilized. Conversely, if treatment efficacy of DBS is compared to sham, then HAMD-6 may be better suited[39]. An optimal study design for future trials should consider methods of blinding patients, where patients are not able to "guess" whether DBS is on or off, as reported in the study by Raymaekers et al [23]. Future studies should compare DBS to a specific type of conventional treatment for TRD such as lithium or ECT.

Advances in target-localization methodologies, from conventional anatomical MRI to individualized tractography, indicate that DBS efficacy in TRD is contingent upon both the selected anatomical target and the fiber bundle. Consequently, a singular "one-target-fits-all" is unlikely to optimize therapeutic outcomes across the heterogeneous TRD population. Future clinical trials should therefore consider personalized targets, rather than restricting DBS to a single region such as SCC, to enhance the efficacy and reproducibility of DBS interventions in TRD.

Conclusion

The results obtained indicate that DBS, when performed correctly at the correct target in the correctly identified patient, is potentially superior to sham. DBS indicated greater efficacy compared to sham in clinical trial phase and long-term follow-up phase. The pooled response rate after long-term follow-up was found to be 56%.

Statements and Declarations

Conflicts of interest

The authors declare that they have no competing interests.

Funding

No funding has been provided.

Author contributions

Study design: All authors. Screening of literature: AS, HA and VA. Data extraction and statistical analysis: AS and HA. Manuscript first draft: AS and HA. Manuscript editing and reviewing: AS, HA, CB, and TM. All authors have read and approved the final revised version of this manuscript.

Ethics approval and consent to participate

Not applicable.

Data sharing

Upon request additional data is available.

Acknowledgements

The authors wish to thank Professor Helen S. Mayberg, Icahn School of Medicine at Mount Sinai, New York, for providing unpublished data regarding patients with treatment-resistant unipolar depression. Dr. Mohtadin Hashemi is acknowledged for giving feedback on the final draft of the paper.

References

1. ^ *James SL, et al. (2018). "Global, Regional, and National Incidence, Prevalence, and Years Lived With Disability for 354 Diseases and Injuries for 195 Countries and Territories, 1990-2017: A Systematic Analysis for the Global Burden of Disease Study 2017." Lancet.* **392**:*1789–1858.*

2. ^Sander J, Moessner M, Bauer S (2021). "Depression, Anxiety and Eating Disorder-Related Impairment: Moderators in Female Adolescents and Young Adults." Int J Environ Res Public Health. **18**:1–12.
3. ^Nutt DJ, Wilson S, Paterson L (2008). "Sleep Disorders as Core Symptoms of Depression." Dialogues Clin Neurosci. **10**:329–336.
4. ^Otte C, et al. (2016). "Major Depressive Disorder." 2.
5. a, bTanner JA, et al. (2020). "Economic Burden of Depression and Associated Resource Use in Manitoba, Canada." Can J Psychiatry. **65**:338–346.
6. a, b, c, dGrover S, Gautam S, Jain A, Gautam M, Vahia V (2017). "Clinical Practice Guidelines for the Management of Depression." Indian J Psychiatry. **59**(5):34–50. doi:10.4103/0019-5545.196973.
7. a, bThase ME, Rush AJ (1997). "When at First You Don't Succeed: Sequential Strategies for Antidepressant Nonresponders." J Clin Psychiatry. **58**.
8. ^Rush AJ, et al. (2006). "Acute and Longer-Term Outcomes in Depressed Outpatients Requiring One or Several Treatment Steps: A STAR*D Report." Am J Psychiatry. **163**:1905–1917.
9. ^Haq AU, Sitzmann AF, Goldman ML, Maixner DF, Mickey BJ (2015). "Response of Depression to Electroconvulsive Therapy: A Meta-Analysis of Clinical Predictors." J Clin Psychiatry. **76**:1374–1384.
10. ^Heijnen WT, Birkenhäger TK, Wierdsma AI, Van Den Broek WW (2010). "Antidepressant Pharmacotherapy Failure and Response to Subsequent Electroconvulsive Therapy: A Meta-Analysis." J Clin Psychopharmacol. **30**:616–619.
11. ^Kellner CH, Goldberg JF, Briggs MC, Liebman LS, Ahle GM (2013). "Somatic Treatments for Severe Bipolar Disorder." Lancet. **382**(9891):505–506. doi:10.1016/s0140-6736(13)61712-2.
12. ^Tokutsu Y, et al. (2013). "Follow-Up Study on Electroconvulsive Therapy in Treatment-Resistant Depressed Patients After Remission: A Chart Review." Clin Psychopharmacol Neurosci. **11**:34–38.
13. ^Dar H, et al. (2023). "Effectiveness of Electroconvulsive Therapy for Preventing Relapse and Recurrence of Depression in Adults With Major Depressive Disorder: An Updated Meta-Analysis of Randomized Clinical Trials." Cureus. doi:10.7759/cureus.35683.
14. ^Chiken S, Nambu A (2016). "Mechanism of Deep Brain Stimulation: Inhibition, Excitation, or Disruption?" Neuroscientist. **22**(3):313–322. doi:10.1177/1073858415581986.
15. ^Mayberg HS, et al. (2005). "Deep Brain Stimulation for Treatment-Resistant Depression." Neuron. **45**:651–660.
16. a, b, c, d, eHitti FL, Yang AI, Cristancho MA, Baltuch GH (2020). "Deep Brain Stimulation Is Effective for Treatment-Resistant Depression: A Meta-Analysis and Meta-Regression." J Clin Med. **9**:2796.
17. ^Millet B, et al. (2014). "Limbic Versus Cognitive Target for Deep Brain Stimulation in Treatment-Resistant Depression: Accumbens More Promising Than Caudate." Eur Neuropsychopharmacol. **24**:1229–1239.
18. a, bBergfeld IO, et al. (2016). "Deep Brain Stimulation of the Ventral Anterior Limb of the Internal Capsule for Treatment-Resistant Depression A Randomized Clinical Trial." JAMA Psychiatry. **73**:456–464.
19. a, b, cDougherty DD, et al. (2015). "A Randomized Sham-Controlled Trial of Deep Brain Stimulation of the Ventral Capsule/Ventral Striatum for Chronic Treatment-Resistant Depression." Biol Psychiatry. **78**:240–248.
20. ^Dersimonian R, Laird N (1986). "Meta-Analysis in Clinical Trials*."
21. a, bTeam RC (2025). "R: A Language and Environment for Statistical Computing." https://www.R-project.org/.
22. ^Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2010). "A Basic Introduction to Fixed-Effect and Random-Effects Models for Meta-Analysis." Res Synth Methods. **1**:97–111.
23. a, b, c, dRaymaekers S, Luyten L, Bervoets C, Gabriëls L, Nuttin B (2017). "Deep Brain Stimulation for Treatment-Resistant Major Depressive Disorder: A Comparison of Two Targets and Long-Term Follow-Up." Transl Psychiatry. **7**:e1251.
24. ^Duval S, Tweedie R (2000). "Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis." Biometrics. **56**:455–463.
25. ^Coenen VA, et al. (2019). "Superolateral Medial Forebrain Bundle Deep Brain Stimulation in Major Depression: A Gateway Trial." Neuropsychopharmacology. **44**:1224–1232.
26. a, bFenoy AJ, et al. (2018). "A Longitudinal Study on Deep Brain Stimulation of the Medial Forebrain Bundle for Treatment-Resistant Depression." Transl Psychiatry. **8**.
27. a, b, cFenoy AJ, et al. (2022). "Deep Brain Stimulation of the "Medial Forebrain Bundle": Sustained Efficacy of Antidepressant Effect Over Years." Mol Psychiatry. **27**:2546–2553.

28. ^Holtzheimer PE, et al. (2012). "Subcallosal Cingulate Deep Brain Stimulation for Treatment-Resistant Unipolar and Bipolar Depression." Arch Gen Psychiatry. **69**:150–158.

29. a, bHoltzheimer PE, et al. (2017). "Subcallosal Cingulate Deep Brain Stimulation for Treatment-Resistant Depression: A Multisite, Randomised, Sham-Controlled Trial." Lancet Psychiatry. **4**:839–849.

30. a, bMerkl A, et al. (2017). "Deep Brain Stimulation of the Subcallosal Cingulate Gyrus in Patients With Treatment-Resistant Depression: A Double-Blinded Randomized Controlled Study and Long-Term Follow-Up in Eight Patients." J Affect Disord. **227**:521–529.

31. a, bPuigdemont D, et al. (2015). "A Randomized Double-Blind Crossover Trial of Deep Brain Stimulation of the Subcallosal Cingulate Gyrus in Patients With Treatment-Resistant Depression: A Pilot Study of Relapse Prevention." J Psychiatry Neurosci. **40**:224–231.

32. ^Ramasubbu R, Anderson S, Haffenden A, Chavda S, Kiss ZHT (2013). "Double-Blind Optimization of Subcallosal Cingulate Deep Brain Stimulation for Treatment-Resistant Depression: A Pilot Study." J Psychiatry Neurosci. **38**:325–332.

33. a, bMorishita T, Fayad SM, Higuchi M, Nestor KA, Foote KD (2014). "Deep Brain Stimulation for Treatment-Resistant Depression: Systematic Review of Clinical Outcomes." Neurotherapeutics. **11**(3):475–484. doi:10.1007/s13311-014-0282-1.

34. a, b, cBerlim MT, McGirr A, Van Den Eynde F, Fleck MPA, Giacobbe P (2014). "Effectiveness and Acceptability of Deep Brain Stimulation (DBS) of the Subgenual Cingulate Cortex for Treatment-Resistant Depression: A Systematic Review and Exploratory Meta-Analysis." J Affect Disord. **159**:31–38.

35. a, bZhou C, et al. (2018). "A Systematic Review and Meta-Analysis of Deep Brain Stimulation in Treatment-Resistant Depression." Prog Neuropsychopharmacol Biol Psychiatry. **82**:224–232.

36. ^Cohen J (1988). Statistical Power Analysis for the Behavioral Sciences. New Jersey: Lawrence Erlbaum Associates.

37. a, bDandekar MP, Fenoy AJ, Carvalho AF, Soares JC, Quevedo J (2018). "Deep Brain Stimulation for Treatment-Resistant Depression: An Integrative Review of Preclinical and Clinical Findings and Translational Implications." Mol Psychiatry. **23**:1094–1112.

38. ^Rabin JS, et al. (2020). "Neuromodulation for Major Depressive Disorder: Innovative Measures to Capture Efficacy and Outcomes." Lancet Psychiatry. **7**(12):1075–1080. doi:10.1016/S2215-0366(20)30187-5.

39. a, b, c, dCarrozzino D, Patierno C, Fava GA, Guidi J (2020). "The Hamilton Rating Scales for Depression: A Critical Review of Clinimetric Properties of Different Versions." Psychother Psychosom. **89**:133–150.

40. ^Bagby RM, Ryder AG, Schuller DR, Marshall MB (2004). "The Hamilton Depression Rating Scale: Has the Gold Standard Become a Lead Weight?" Am J Psychiatry. **161**:2163–2177.

41. ^Santen G, Danhof M, Pasqua O Della (2009). "Sensitivity of the Montgomery Asberg Depression Rating Scale to Response and Its Consequences for the Assessment of Efficacy." J Psychiatr Res. **43**:1049–1056.

42. ^Ruhé HG, Van Rooijen G, Spijker J, Peeters FPML, Schene AH (2012). "Staging Methods for Treatment Resistant Depression. A Systematic Review." J Affect Disord. **137**:35–45.

43. ^Schosser A, et al. (2012). "European Group for the Study of Resistant Depression (GSRD) - Where Have We Gone So Far: Review of Clinical and Genetic Findings." Eur Neuropsychopharmacol. **22**:453–468.

Supplementary data: available at https://doi.org/10.32388/47TW5N

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.