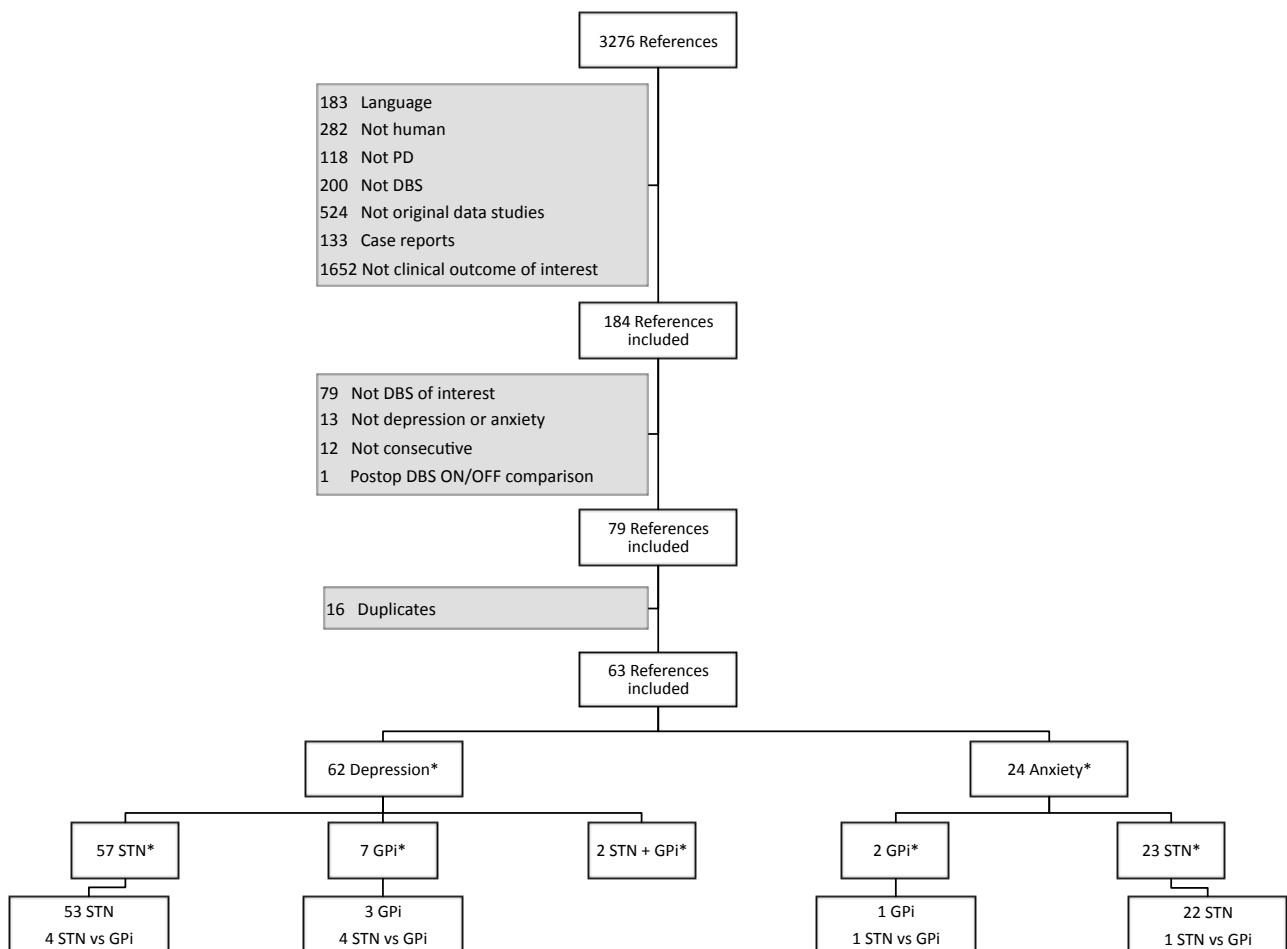


## Supplementary online material

### Appendix 1. Systematic review flow-chart



Systematic review approach is outlined with respective proportion of excluded references attributed to each criterion. Literature distribution by psychiatric outcome and surgical target is also presented.

\* The same reference might be contained in more than one group.

Abbreviations: PD: Parkinson's Disease; DBS: Deep Brain Stimulation; STN: Subthalamic Nucleus; GPi: Globus Pallidus internum; STN + GPi: studies with data not discriminated regarding both targets; STN vs GPi: studies with comparison between both targets.

## Appendix 2. Included references details

[illegible]

Psychometric instruments used are highlighted with “x”. Stimulation target is marked with “✓” in “STN” and/or “GPI” columns. Follow-up studies have “✓” in the respective column and

comparison studies are codified by 1 to 4 so different comparators can be distinguished. Grey shading denotes references excluded from analysis (non-comparable data).

Notes: 1 to 4 corresponds to comparators coding; 1: healthy control group, 2: medical treatment control group, 3 eligible for surgery control group and 4: GPi comparison group; \* references with STN and GPi data not discriminated: no further analysis; (a) original data is pre M(SD) and post M(SD): changeM calculated as (postM - preM); changeSD calculated as  $\sqrt{[(\text{preSD}^2 + \text{posSD}^2 - 2 \times \text{preSD} \times \text{posSD}) / n]}$ ; (b) original data is pre M(SD) and change M(SD); (c) original data not quantitative/comparable (percentage of patients): no further analysis; (d) original data not quantitative/comparable (qualitative description): no further analysis; (e) within each period of time, the longest follow-up was selected for the analysis; (f) "x to y months" type follow-up: y months assumed; (g) original data reported by groups: separately considered for the analysis; (h) original data reported on total sample and by groups: total sample considered; (i) SD calculated from SE as  $(\text{SE} \times \sqrt{n})$ ; (j) original data reported in on and off state: only on considered; (k) original data not quantitative/comparable (no dispersion measure): no further analysis; (l) mean (SD) assumed; (m) SD calculated from 95% CI as  $[(\text{upper limit} - \text{lower limit}) / 3.92] \times \sqrt{n}$ ; (n) original data reported individually: preM(SD) and changeM(SD) calculated; (o) graphical data; (p) intention-to-treat analysis; (q) original data not quantitative/comparable (percentage of change): no further analysis; (A) cognitive outcomes compared with control group; depression and anxiety assessed only in patients; so, follow-up STN-DBS study design assumed; (B) "positive change scores indicate clinical improvement; data are (...) mean (SD) (...) for changes between baseline (before DBS) and 6 months": - changeM assumed; (C) HDRS not consecutively assessed: "depression was evaluated (...) using the Self-Rating Depression Scale (...); every patient whose SDS score showed a mild depression, or more, was evaluated again (...) using the Hamilton Depression Scale"; (D) partial duplicates: BDI, POMS-d, STAI-s, STAI-t, SCL-90-R-d and SCL-90-R-a data from Kaiser, 2008; BRMES and HAMA data from Kalteis, 2006; (E) partial duplicates:

3 years follow-up data from Zibetti, 2009; 9 years follow-up data from Zibetti, 2011 (and the respective preoperative data for each one); (F) data from n=20 (whole sample) and from n=9 (18 months follow up sample); evaluation moments at 3, 6, 12 and 18 months; n=20 preoperative data considered for short-term follow-up analysis; n=9 preoperative data considered for mid-term follow-up analysis; (G) stimulation device was turned on 4 weeks after the surgery: postoperative moments converted to post-DBS moments by subtracting 1 month; (H) "depression severity index" = "accumulative scores of each item"/"maximum scores of the scale": mean x 80 and SD x 80 assumed; (I) "the assessments took place (...) 12 months (...) later, with the exception of the cognitive status, which was controlled 3 months after surgery"; "outcome measures" = "motor function" + "cognitive status" + "psychiatric history" + "mood and behavioral modifications: ardouin scale" + "acute non-motor fluctuations": 1 year follow up assumed to mood evaluation; (J) results separated by groups "identical", "ameliorated" and "aggravated": not comparable with other studies; (K) Partial duplicate: 4 groups: STN versus GPI and Paris versus Grenoble: GPI in Grenoble, GPI versus STN comparison in Grenoble, STN in Paris, GPI in Paris and GPI versus STN comparison in Paris included; STN in Grenoble duplicated; (L) 4 groups: STN versus GPI and Paris versus Grenoble: only 57 in a total of 62 patients performed BDI assessment and the distribution by groups was not indicated: total n assumed for each group. (M) staged DBS; evaluation times were "1 month before first surgery, 2 months following first surgery (unilateral), and 3 months following second surgery (bilateral)": 3 months follow-up assumed; (N) "Test-retest interval was about 3 months between baseline and post-unilateral electrode placement evaluation, and 4 months between post-unilateral and post-bilateral electrode placement evaluations. This occurred with the exception of one patient who on separate occasions had the lead and pulse generator repositioned following bilateral operation, resulting in a 22-month lapse between neuropsychological assessments after first and second DBS electrode placement.": global 3 months follow-up assumed; (O) pre- and postoperative evaluations performed in patients group; only 1 evaluation in control group: postoperative cross-sectional analysis assumed.

Test-retest coefficient (r) was 0,66 for BAI<sup>83</sup>, 0,64 (short term) and 0,75 (mid- and long- term) for BDI<sup>81</sup>, 0,79 for BSI-a<sup>85</sup>, 0,84 for BSI-d<sup>85</sup>, 0,94 for GDS<sup>84</sup>, 0,98 for HAD-a<sup>88</sup>, 0,99 for HAD-d<sup>88</sup>, 0,87 for HDRS<sup>79</sup>, 0,56 for MADRS<sup>80</sup>, 0,4 for STAI-s<sup>90</sup>, 0,86 for STAI-t<sup>90</sup> and 0,651 for UPDRS I,3<sup>86</sup>. 0,98 assumed for BRMES<sup>87</sup>. 0,75 assumed for POMS-d<sup>89</sup>. Conservative value of 0,56 was assumed for SCL-90-R-d and Zung-d. Conservative value of 0,4 was assumed for AMDP-AT, BAS, HAMA, SCL-90-R-a and Zung-a.

Abbreviations: in “follow-up” column, w, m and y refers to weeks, months and years, respectively; pre: preoperative data; post: postoperative data; M: mean; SD: standard deviation; SE: standard error; 95%CI: 95% confidence interval; r: test-retest correlation coefficient; n: sample size; "change" refers to the postop - preop temporal change; "difference" refers to the STN - comparison group difference; AMDP-AT: association for methodology and documentation In psychiatry, anxiety part; Ardouin-a and Ardouin-d: “anxiety” and “depressive mood” items of the Ardouin scale, respectively; BAI: Beck anxiety inventory; BAS: brief scale for anxiety; BDI: Beck depression inventory; BRMES: Bech-Rafaelsen Melancholia Scale; BSI-a and BSI-d: anxiety and depression scales of the brief symptom inventory, respectively; GDS: geriatric depression scale; GPi: globus pallidus, pars interna; HAD-a and HAD-d: anxiety and depression parts of the hospital anxiety and depression scale, respectively; HAMA: Hamilton anxiety scale; HDRS: Hamilton depression rating scale; IOWA-a and IOWA-d: anxiety and depression parts of the IOWA scales of personality change, respectively; MADRS: Montgomery-Asberg depression rating scale; MINI-a and MINI-d: “general anxiety”/”anxiety disorders” and “major depression episode/disorder” items of the “mini international neuropsychiatric interview, respectively; NMSQuest-a and NMSQuest-d: items “anxiety” and “feeling sad” of the non motor symptom questionnaire, respectively; POMS-d: profile of mood states, depression domain; SCL-90-R-a and SCL-90-R-d: anxiety and depression domains of the symptom checklist-90-revised; STAI-s and STAI-t: state and trait (respectively) anxiety inventory; STN: subthalamic nucleus; UPDRS I,3: unified parkinson’s disease rating scale,

part I, item 3 “depression”; Zung-a and Zung-d: Zung self-rating anxiety and depression scales, respectively. Country abbreviations according to ISO 3166-1 decoding table.