Towards individualized deep brain stimulation: A stereoencephalography-based workflow for unbiased neurostimulation target identification

Running Title: Workflow for stimulation target discovery

Jeremy Saal MS*[a,b], Kelly Kadlec PhD*[a,c,d], Anusha B. Allawala PhD[a,b], Lucille Johnston BS[a,b], Ryan B. Leriche BS[a,b], Ritwik Vatsyayan PhD[a,b], Yiyuan Han PhD[a,b], Audrey Kist PhD[a,d], Tommaso Di Ianni PhD[a,d,e], Heather E. Dawes PhD[a,b], Edward F. Chang MD[a,b], A Moses Lee MD PhD[a,d], Andrew D. Krystal MD[a,d], Khaled Moussawi MD PhD[a,c,d], Prasad Shirvalkar MD PhD[a,b,c,f], Kristin K. Sellers PhD[a,b]

a. UCSF Weill Institute for Neurosciences, University of California San Francisco, San Francisco, CA, USA.
b. Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA, USA.
c. Department of Neurology, University of California, San Francisco, San Francisco, CA, USA.
d. Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, San Francisco, CA, USA.
e. Department of Radiology and Biomedical Imaging University of California, San Francisco, CA, USA.
f. UCSF Department of Anesthesiology and Perioperative Care, Division of Pain Medicine, University of California San Francisco, San Francisco, CA, USA.
* These authors contributed equally

Sources of Financial Support: This work was supported by the Ray and Dagmar Dolby Family Fund through the Department of Psychiatry at UCSF, Foundation for OCD Research, the National Institute of Mental Health under Award Number R21MH130914, the National Institute of Neurological Disorders and Stroke of the National Institutes of Health under Award Number UH3NS123310, the National Institute of Health HEAL Initiative Grant UH3NS115631, and the UCSF Weill Institute for Neurosciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authorship Statements:
JS, KK, ABA, LJ, RBL, RV, YH, AK, TDI, HED, EFC, AML, ADK, KM, PS, and KKS designed and conducted the studies, including patient recruitment, data collection, data analysis, and interpretation of the data. JS, KK, RV, YH, KKS prepared the manuscript draft with intellectual input from ABA, LJ, AK, TDI, HED, AML, ADK, KM, and PS. All authors approved the final manuscript

Conflict of Interest Statement: All authors have completed required COI forms

Corresponding author:
Kristin K. Sellers, PhD
Weill Neurosciences Building, Box 3128
1651 4th Street, Room 671B
San Francisco, CA 94143
kristin.sellers@ucsf.edu
Phone: 630.204.9534

Abstract:

Objectives: Deep brain stimulation (DBS) is increasingly being used to treat a variety of neuropsychiatric conditions, many of which exhibit idiosyncratic symptom presentations and neural correlates across individuals. Thus, we have utilized inpatient stereoelectroencephalography (sEEG) to identify personalized therapeutic stimulation sites for chronic implantation of DBS. Informed by our experience, we have developed a statistics-driven framework for unbiased stimulation testing to identify therapeutic targets.

Materials and Methods: Fourteen participants (major depressive disorder = 6, chronic pain = 6, obsessive-compulsive disorder = 2) underwent inpatient testing using sEEG and symptom monitoring to identify personalized stimulation targets for subsequent DBS implantation. We present a structured approach to this sEEG testing, integrating a Stimulation Testing Decision Tree with power analysis and effect size considerations to inform adequately powered results to detect therapeutic stimulation sites with statistical rigor.

Results: Effect sizes (Cohen's d) of stimulation-induced symptom score changes ranged from -1.59 to +2.59. The standard deviation of sham trial responses was strongly correlated with the standard deviation of stimulation responses ($r = 0.86$, $p < 0.001$), and thus could be used to estimate the variability of stimulation responses for power analysis calculations. We show that 12-15 sham trials were needed to robustly estimate sham variability. Power analysis (using a paired-t test) showed that for effect sizes ≥ 1.1, approximately 10 trials should be used per stimulation site for sufficiently powered results.

Conclusions: The workflow presented is adaptable to any indication and is specifically designed to overcome key challenges experienced during stimulation site testing. Through

incorporating sham trials, effect size calculations, and tolerability testing, the described approach can be used to identify personalized, unbiased, and clinically efficacious stimulation sites.

Key words: Stereoencephalography, deep brain stimulation, neurostimulation, personalized targets, clinical workflow

Introduction

Deep brain stimulation (DBS) is emerging as a promising therapeutic modality for a wide range of indications, particularly neuropsychiatric disorders that involve underlying network dysfunctions. Unique perturbations in network function across individuals have led to interest in testing stimulation across brain regions to identify personalized therapeutic sites. To achieve this, our center at the University of California, San Francisco and others utilize inpatient stereoelectroencephalography (sEEG) to explore stimulation across the disease network, aiming to identify personalized therapeutic targets for chronic implantation with DBS (1–6). Throughout, we use the overarching term “DBS” to refer to both DBS devices and responsive neurostimulation using the NeuroPace RNS System. sEEG inpatient monitoring is the established standard of care for complex cases of refractory epilepsy (7,8). We leverage the existing surgical procedures and specialty inpatient units to care for patients with temporarily implanted sEEG leads and have conducted sEEG monitoring stays lasting approximately 10 days with 14 participants across three clinical indications (major depressive disorder [MDD], chronic pain [CP], and obsessive-compulsive disorder [OCD]). Informed by our experience, we have developed a statistics-driven framework for unbiased stimulation testing to identify therapeutic targets.

DBS for complex neuropsychiatric disorders has had variable clinical efficacy across individuals and indications. DBS for MDD has primarily targeted the subcallosal cingulate (9) and the ventral capsule/ventral striatum (VC/VS) (10). While smaller open-label studies and case-reports across anatomical targets have shown initial success (11), results from randomized sham-controlled clinical trials have shown considerable heterogeneity (12). Investigations into DBS for chronic pain date back to the 1950s, revealing potential targets including the anterior cingulate cortex, ventral posterolateral and ventral posteromedial nuclei of the thalamus, and the periventricular grey (PVG) (13,14). Clinical outcomes of DBS for CP are promising, with some studies reporting that over half of patients experience more than 50% pain

reduction within 12 months of implantation (15). However, no published DBS trial for CP to date has shown superior pain relief to placebo (16). In the context of OCD, the FDA granted a Humanitarian Device Exemption in 2009 for DBS targeting the anterior limb of the internal capsule. Since then, stimulation has also been targeted to the VC/VS, bed nucleus of the stria terminalis, and subthalamic nucleus (17,18). Yet a third of patients show no response (19), underscoring the need for individualized targeting. Our staged sEEG-DBS approach aims to improve clinical efficacy by personalizing DBS targets.

Current DBS devices support one to four leads, limiting anatomical coverage. Though revision surgeries to reposition leads are possible, they are costly and increase surgical risk. Hence, optimal implantation site (location of stimulation electrode contacts) selection is crucial. Towards this objective, our testing framework is designed with three goals in mind: (1) standardize stimulation procedures for rigorous, unbiased results; (2) minimize risk and burden; and (3) identify at least one robust therapeutic target.

To address these goals, we present here a Stimulation Testing Decision Tree designed to facilitate the identification of personalized stimulation sites. Importantly, the Decision Tree incorporates power analysis and effect size considerations based on data from participants who underwent stimulation testing with sEEG in 3 indications (MDD, CP, OCD), to inform the design of stimulation testing which will be adequately powered to detect therapeutic stimulation sites with statistical rigor. We propose this testing framework to avoid pitfalls and obstacles that we have identified during prior sEEG testing periods, including:

1. Trying to (over) optimize one site of stimulation: Without a principled testing approach, upon observing therapeutic benefit with stimulation, it can be tempting to adjust stimulation parameters to try and ‘optimize’ stimulation at that site. However, this is often done at the expense of obtaining sufficient trials with stable parameters to statistically

power a determination of benefit, testing other sites, or conducting sufficient sham-trial testing.

2. Over-reliance or under-reliance on quantitative metrics to assess therapeutic benefit: Even with training on self-report measures, participants adopt idiosyncratic behavior in responding to symptom surveys, such as only responding within a narrow range of the scale. Thus, over-reliance on strict quantitative cutoffs may not fully capture beneficial stimulation. Conversely, basing surgical implant decisions entirely on clinician observations may suffer from personal bias and inter-rater reliability concerns.
3. Weighing stimulation benefits and side effects: Even if stimulation is highly beneficial for symptom relief, if it is accompanied by concerning side effects, other sites should be pursued for benefit. Stimulation-induced side effects may prevent participant blinding during randomized controlled trials. Stimulation side effects span a wide range of experiences, including sweaty hands, buzzing or tingling, bladder spasms, dissociation, vertigo, induced smiling, and others.

Materials and Methods

Study Participants and Clinical Environment

As part of respective clinical trials, 14 participants were implanted with sEEG for approximately 10 days of inpatient monitoring to identify brain regions in which acute stimulation led to reduction in primary indication symptoms (MDD: n = 6, CP: n = 6, OCD: n = 2). PMT (PMT Corporation, Chanhassen, MN), AdTech (AdTech Medical, Oak Creek, WI), or DIXI (DIXI Medical, Marchaux Chaudefontaine, France) electrodes were implanted bilaterally in regions implicated in the primary indication. Anatomical localization of electrode contacts was conducted by co-registration of preoperative T1 MRI images and postoperative CT scan and automated parcellation and manual verification of the labeling. Contacts within defined regions of interest were selected for subsequent stimulation testing. Electrode tails were connected to cables that

interfaced with a 256-channel Nihon Kohden neurophysiology recording and stimulation system. Charge-balanced, biphasic, bipolar stimulation was delivered between one anode and one cathode. The duration of stimulation used for testing ranged across indications based on expected and observed modulation of symptoms (10s to 20 minutes of continuous stimulation).

Participants provided frequent symptom reports via visual analog scale (VAS) ratings on a tablet computer. VAS sliders were horizontal and ranged from 0 (no presence of that symptom) to 100 (highest presence of that symptom). VAS symptoms were indication-specific (MDD: depression; CP: pain intensity; OCD: obsessions, compulsions, and OCD-related distress [scores were summed to create an OCD composite score]). Scores were collected before stimulation trials (baseline), during stimulation or sham trials, and during washout. These data were collected and managed using REDCap electronic data capture tools hosted at UCSF (20,21). REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for data integration and interoperability with external sources. Change in VAS score (score during stimulation/sham minus score during the immediately preceding baseline or washout period, termed ‘change score’) was used to quantitatively assess for symptom changes induced by stimulation. We focus our analysis on VAS survey responses because of its utility for momentary symptom tracking. Furthermore, VAS change scores generalize across indications more readily than other metrics. All study activities were approved by the Committee on Human Research at the University of California, San Francisco and covered under FDA Investigational Device Exemptions for each indication (MDD: NCT04004169, CP: NCT04144972, OCD: NCT06347978).

Statistical Analysis

To measure effect size, Cohen's d was calculated using sham and verum stimulation data collected from the 14 participants in the cohort, according to the equation in Fig 2C. Stimulation sites with ≥ 3 trials were included in analyses. For each participant, the "best" stimulation site was identified as the site with the largest Cohen's d compared to sham. Hedge's G was also calculated as an alternative effect size metric. Standard deviations (SD) across trials were computed separately for sham and the best site for each participant. To verify whether variability in sham trials can be used to estimate verum stimulation response SDs (and thus can be used to inform power analysis), we calculate the Pearson correlation between sham and stimulation response SDs. We used linear regression to estimate how sham trial variability scaled with best-site stimulation variability.

Number of Trials Calculation

To estimate the required sample size at different clinically relevant effect sizes (0.8 to 1.7) to achieve sufficiently powered results (power = 0.8), we used the open-source computing tool G*Power (22). Given the paired design proposed for future studies, we first conducted this power analysis using a paired one-tailed t-test with a power of 0.8 and alpha = 0.05 (one-tailed because the desired effect is in one direction only). Bonferroni multiple comparisons correction was applied to this calculation to determine the required sample size when multiple regions were tested (e.g. alpha = 0.05 for 1 region; alpha = 0.005 for 10 regions). A similar procedure was followed using unpaired t-tests and paired Wilcoxon signed rank to assess sample sizes for unpaired and nonparametric distributions, respectively.

Stability of sham standard deviation across trials

To inform how many preliminary efficacy trials are needed to estimate response variability, we sought to estimate the minimum number of stimulation trials required for a stable

estimation of the full distribution from existing data. For each participant, we bootstrapped the sham standard deviations, starting with two trials and increasing up to the total number of sham trials (10,000 permutations). For each sample size, we calculated the proportion of estimates that fell within 5% of the true standard deviation of the full dataset. Lastly, to test if sham-trial standard deviation remained stable across the sEEG testing period, we compared the standard deviations of the first and second halves of temporally ordered sham trials using a paired t-test.

Results

Stimulation Testing Decision Tree for Identifying a Therapeutic Stimulation Site

Using data collected from sEEG monitoring phases with 14 participants, we formalized a Stimulation Testing Decision Tree (Fig 1). The decision tree can be followed for testing an individual participant and is designed to optimize decision-making based on positive evidence while taking into consideration key points of identified risk. The key parameter to 'map' during inpatient testing is the therapeutic stimulation site. Stimulation intensity is adjusted as needed during testing. While we recommend keeping the stimulation frequency consistent, it can also be changed and included as a different site for statistical purposes. We recommend holding pulse-width constant to reduce the parameter space and, thus, increase statistical power. The different color-coded sections of the decision tree reflect key considerations to enable unbiased decision making about where to implant chronic DBS for therapeutic benefit.

1. **Evidence for therapeutic efficacy** (Fig 1, blue):

   The first seven steps described below occur during the inpatient monitoring phase and initial contact verification and selection.

   a. *Safety testing*: All stimulation sites should first undergo safety testing. We recommend 1s and then 3s at 1, 3, and 6mA (if using electrodes with 2mm

contacts and 0.8mm diameter). If afterdischarges are observed, see ‘Sufficient safe dynamic range’ section below.

b. *Tolerability testing*: Ideally, stimulation does not elicit side effects. If any somatic, sensory, or behavioral side effects are reported or observed with stimulation testing, see ‘Tolerability and side effects of stimulation’ section below.

c. *Spontaneous vs. provoked symptoms*: Symptoms must be present to effectively test for improvement with stimulation. While it remains poorly understood if there is a shared neurobiological substrate for elicited vs. spontaneous symptoms, the limited timeframe and environment of inpatient testing may necessitate that symptoms are provoked to allow stimulation testing. Symptom provocation should be thoroughly discussed with participants prior to initiation, including allowable or non-allowable topics of discussion or tasks. To date, our experience is that stimulation shown to be acutely beneficial in the inpatient context of symptom provocation for MDD and OCD is also beneficial during chronic DBS in the outpatient setting. If symptoms are expected to remain low during sham-controlled testing, provocation may need to be incorporated, as for example in Seilheimer et al., 2025 (23).

d. *Preliminary efficacy testing to estimate effect size*: The minimum effect size that corresponds to clinical improvement can be estimated using a change score value with clinical significance along with the standard deviation of the change scores from sham trials collected during preliminary efficacy trials. In our experience, variability in symptom score reporting is similar between sham and verum stimulation trials ($r = 0.86$, $p < 0.001$). During preliminary efficacy testing, a higher number of sham trials relative to verum stimulation trials should be included for accurate SD estimates (see Table 1 for suggested trial design).

e. *Determine the number of sites and number of trials per site for testing*: The estimated effect size can then be used to determine the number of trials required per stimulation site being tested.

f. *Select stimulation sites*: If there are too many sites remaining after safety and tolerability testing to feasibly acquire the necessary number of trials per site, the number of sites to be tested should be reduced. This can be done using methods appropriate for the particular indication or participant, for example examining tractography, using evoked potentials to evaluate connectivity, performing further stimulation testing to identify sites with strong/rapid behavioral responses, or using prior literature.

g. *Sham-controlled testing*: Stimulation parameters should use a stimulation intensity that was determined to be safe and tolerable during safety testing.

h. *Calculate symptom change scores*: A symptom score will be taken immediately preceding each stimulation or sham trial and in response to each stimulation/sham trial in order to assess for change in symptoms.

i. *Statistical tests*: Either a paired t-test or a paired signed rank test can be used to compare the response scores collected within each stimulation block to determine the therapeutic effect of a stimulation site compared with sham.

j. *Results interpretation*: If the statistical tests demonstrate significance for a site compared to sham and this is consistent with clinical observation and participant verbal report, the site can be considered a potential therapeutic target for chronic implantation. If multiple sites are identified as being therapeutic, the most effective site can be determined by ranking therapeutic target change scores. If the statistical test is not significant, other metrics (e.g. participant verbal reports, clinician scales) should be investigated for evidence of therapeutic benefit. If no

clinical benefit is captured by these metrics, no therapeutic site has been identified.

2. **Sufficient safe dynamic range** (Fig 1, green): The presence of afterdischarges with high amplitude stimulation should warrant extra vigilance at that site. However, safe stimulation may be possible at lower amplitudes (e.g. afterdischarges seen at 6mA do not preclude safe stimulation at 2mA). The threshold for afterdischarges is dynamic, so further stimulation with lower amplitude at a site that exhibited afterdischarges should be undertaken with caution.
3. **Tolerability and side effects of stimulation** (Fig 1, orange): Stimulation-induced side effects can generally be categorized as unpleasant or neutral/pleasant. For example, sensory thalamus stimulation in patients with chronic pain has been shown to induce pleasant paresthesias (24) and PVG stimulation has coincided with a warm sensation (25,26). Unpleasant side effects must be avoided, and the site only remains viable if the side effect is not present with lowered stimulation amplitude. The threshold for side effects is dynamic, so sufficient amplitude range without side effects must be present during acute testing. If participant blinding is necessary, as is the case for many clinical trials, sites should only be advanced if amplitude can be lowered to the point of no perceptible side effects, even if side effects are neutral or positive. If blinding is not needed, clinicians/researchers should carefully evaluate together with participants if the presence of neutral or positive side effects would be distracting or maladaptive in daily life.

Empirical determination of sample size

Across all participants and all stimulation sites, effect sizes ranged from −1.59 to +2.59 (Fig 2A). We are interested in the greatest positive effect size within each participant, which indicates an improvement in symptoms due to stimulation (range of 2.59 to 0.11 across the 14

participants).Given the invasive nature of the proposed intervention, we prioritized stimulation sites with large effect sizes. To account for small sample sizes, Hedge's G was also calculated and led to a similar range of effect sizes, ranging from 2.39 to 0.1 (Supplemental Fig 1). Each participant's sham SD versus their best-site SD revealed a strong positive relationship ($r = 0.86$, $p < 0.001$), which supports our proposed use of the SD of sham trials from preliminary efficacy testing to estimate effect size (Fig 2B). A linear regression showed a slope of 1.05 ($p < 0.001$) and an intercept of 3.47, indicating a near one-to-one relationship between sham variability and best-site stimulation variability. Participants with higher sham variability (e.g., those whose sham SD exceeded 20) tended to have equally large best-site variability (e.g., 30+), suggesting trait-level characteristics in the variability of score reporting. The estimated sham and stimulation sites SDs, combined with a predetermined change score representing the minimum clinically significant improvement, can be used to calculate the effect size for each participant using Cohen's d (Fig. 2C). Alternatively, for small sample sizes, Hedge's G can be used to estimate the effect size. This effect size can then be used to determine the number of sites that can be tested and the number of trials needed per stimulation site.

Using the statistical power analysis tool G*power (22), we computed the number of trials required for different effect sizes and numbers of stimulation sites. For effect sizes ≥1.1, around 10 samples per site are needed for statistically powered results (Fig 2D). The same analysis performed for the case of an unpaired t-test shows a substantial increase in the number of samples required per site, emphasizing the importance of a paired design with matched trials for sham and stimulation site (Supplemental Fig 2A). This increase in required samples per site does not occur when using a paired nonparametric test for non-normally distributed data (Supplemental Fig 2B).

We propose a blocked trial design in which each block has at least one sham trial plus any number of verum stimulation sites within a given testing session (Supplemental Table 1). Using existing data, we determined that 12-15 sham trials are needed to robustly estimate the

true standard deviation using bootstrapping (Supplemental Fig 2C), effect size, and subsequently the desired number of trials required for each site. While data from existing participants were not collected using the proposed preliminary efficacy testing, a paired-samples t-test revealed no significant difference in sham response variability between the first and second half of trials ($t(13) = 0.756$, $p = 0.463$), supporting the stability of sham trial variability for estimating stimulation response SD (Supplemental Fig 2D).

Discussion

We propose an empirically derived and generalizable sEEG-based workflow to identify therapeutic sites for chronic implantation of DBS electrodes to treat a variety of neuropsychiatric conditions. Using lessons learned from 14 participants undergoing a sEEG trial period, this workflow involves clear decision-making rationale for both inpatient testing and post-hoc analyses. Our sEEG-based decision tree and workflow offers a prospective method to systematically reduce bias while enhancing both personalization and standardization in chronic DBS implantation site selection (27,28). By leveraging trait-level variability in symptom score reporting, we can estimate effect sizes from sham trials during preliminary efficacy testing and determine the number of trials needed per stimulation site to provide statistically powered results.

Despite its advantages, this workflow has several limitations. It assumes that stimulation effects emerge and subside within minutes, potentially overlooking slower-acting effects. While sham trials and washout periods help mitigate carryover concerns, available time may still be insufficient for complete symptom reversion. Untested parameters at a “ruled-out” site could theoretically outperform the chosen implantation site. The proposed framework does not guarantee the identification of a therapeutic target. Rather, it provides testing guidelines which, based on real participant data, are likely to provide sufficient statistical power to identify a therapeutic target if it exists within the regions tested. In future studies, data-driven methods

based on Bayesian optimization or machine learning models could be integrated with our framework to effectively optimize over a larger parameter space (29–31). Lastly, this framework does not specifically facilitate the assessment of state-dependent effects of stimulation, as testing is aggregated across multiple days and states.

Conclusion

To summarize, we have proposed a data-driven workflow that leverages sEEG-based stimulation to address key challenges in personalizing DBS. The proposed approach allows the identification of therapeutic sites by integrating rigorous trial design, effect size analysis, and sham comparisons. While this approach should be further validated, our framework provides a pragmatic method for optimizing DBS targets and parameters across diverse neuropsychiatric conditions, thus laying the groundwork for more precise and effective neuromodulation therapies.

References

1. Starkweather CK, Sugrue LP, Cajigas I, Speidel B, Krystal AD, Scangos K, Chang EF (2024): Stereoelectroencephalography Electrode Implantation for Inpatient Workup of Treatment-Resistant Depression. *Neurosurgery* 95: 941–948.
2. Shirvalkar P, Sellers KK, Schmitgen A, Prosky J, Joseph I, Starr PA, Chang EF (2020): A Deep Brain Stimulation Trial Period for Treating Chronic Pain. *J Clin Med* 9: 3155.
3. Scangos KW, Khambhati AN, Daly PM, Makhoul GS, Sugrue LP, Zamanian H, *et al.* (2021): Closed-loop neuromodulation in an individual with treatment-resistant depression. *Nat Med* 27: 1696–1700.
4. Sheth SA, Shofty B, Allawala A, Xiao J, Adkinson JA, Mathura RK, *et al.* (2023): Stereo-EEG-guided network modulation for psychiatric disorders: Surgical considerations. *Brain Stimulat* 16: 1792–1798.
5. Allawala A, Bijanki KR, Goodman W, Cohn JF, Viswanathan A, Yoshor D, *et al.* (2021): A Novel Framework for Network-Targeted Neuropsychiatric Deep Brain Stimulation. *Neurosurgery* 89: E116–E121.
6. Sheth SA, Bijanki KR, Metzger B, Allawala A, Pirtle V, Adkinson JA, *et al.* (2022): Deep Brain Stimulation for Depression Informed by Intracranial Recordings. *Biol Psychiatry* 92: 246–251.
7. Theodore WH, Fisher RS (2004): Brain stimulation for epilepsy. *Lancet Neurol* 3: 111–118.
8. Delgado-Garcia G, Frauscher B (2022): Future of Neurology & Technology: Stereoelectroencephalography in Presurgical Epilepsy Evaluation. *Neurology* 98. https://doi.org/10.1212/WNL.0000000000013088
9. Malone DA, Dougherty DD, Rezai AR, Carpenter LL, Friehs GM, Eskandar EN, *et al.* (2009): Deep Brain Stimulation of the Ventral Capsule/Ventral Striatum for Treatment-Resistant Depression. *Biol Psychiatry* 65: 267–275.

10. Lozano AM, Mayberg HS, Giacobbe P, Hamani C, Craddock RC, Kennedy SH (2008): Subcallosal Cingulate Gyrus Deep Brain Stimulation for Treatment-Resistant Depression. *Biol Psychiatry* 64: 461–467.
11. Hitti FL, Cristancho MA, Yang AI, O'Reardon JP, Bhati MT, Baltuch GH (2021): Deep Brain Stimulation of the Ventral Capsule/Ventral Striatum for Treatment-Resistant Depression: A Decade of Clinical Follow-Up. *J Clin Psychiatry* 82. https://doi.org/10.4088/JCP.21m13973
12. Dougherty DD, Rezai AR, Carpenter LL, Howland RH, Bhati MT, O'Reardon JP, *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.
13. Shirvalkar P, Veuthey TL, Dawes HE, Chang EF (2018): Closed-Loop Deep Brain Stimulation for Refractory Chronic Pain. *Front Comput Neurosci* 12: 18.
14. Boccard SGJ, Pereira EAC, Aziz TZ (2015): Deep brain stimulation for chronic pain. *J Clin Neurosci* 22: 1537–1543.
15. Farrell SM, Green A, Aziz T (2018): The Current State of Deep Brain Stimulation for Chronic Pain and Its Context in Other Forms of Neuromodulation. *Brain Sci* 8: 158.
16. Lempka SF, Malone DA, Hu B, Baker KB, Wyant A, Ozinga JG, *et al.* (2017): Randomized clinical trial of deep brain stimulation for poststroke pain. *Ann Neurol* 81: 653–663.
17. De Koning PP, Figee M, Van Den Munckhof P, Schuurman PR, Denys D (2011): Current Status of Deep Brain Stimulation for Obsessive-Compulsive Disorder: A Clinical Review of Different Targets. *Curr Psychiatry Rep* 13: 274–282.
18. Mosley PE, Windels F, Morris J, Coyne T, Marsh R, Giorni A, *et al.* (2021): A randomised, double-blind, sham-controlled trial of deep brain stimulation of the bed nucleus of the stria terminalis for treatment-resistant obsessive-compulsive disorder. *Transl Psychiatry* 11: 190.

19. Gadot R, Najera R, Hirani S, Anand A, Storch E, Goodman WK, *et al.* (2022): Efficacy of deep brain stimulation for treatment-resistant obsessive-compulsive disorder: systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry* 93: 1166–1173.

20. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG (2009): Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 42: 377–381.

21. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, *et al.* (2019): The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform* 95: 103208.

22. Faul F, Erdfelder E, Lang A-G, Buchner A (2007): G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39: 175–191.

23. Seilheimer RL, Qiu L, Rocchio G, Nho Y-H, Campos G, Horn A, *et al.* (2025): Stereo-encephalography-guided multi-lead deep brain stimulation for treatment-refractory obsessive compulsive disorder – study design and individualized surgical targeting approach. *medRxiv.*

24. Pereira EAC, Boccard SG, Linhares P, Chamadoira C, Rosas MJ, Abreu P, *et al.* (2013): Thalamic deep brain stimulation for neuropathic pain after amputation or brachial plexus avulsion. *Neurosurg Focus* 35: E7.

25. Young RF, Chambi VI (1987): Pain relief by electrical stimulation of the periaqueductal and periventricular gray matter: Evidence for a non-opioid mechanism. *J Neurosurg* 66: 364–371.

26. Richardson DE, Akil H (1977): Pain reduction by electrical brain stimulation in man. *J Neurosurg* 47: 178–183.

27. Chang B, Park JJ, Buch VP (2024): Applying normative atlases in deep brain stimulation: a comprehensive review. *Int J Surg* 110: 8037–8044.

28. Hollunder B, Rajamani N, Siddiqi SH, Finke C, Kühn AA, Mayberg HS, *et al.* (2022): Toward personalized medicine in connectomic deep brain stimulation. *Prog Neurobiol* 210: 102211.

29. Stieve BJ, Richner TJ, Krook-Magnuson C, Netoff TI, Krook-Magnuson E (2023): Optimization of closed-loop electrical stimulation enables robust cerebellar-directed seizure control. *Brain* 146: 91–108.

30. Minai Y, Soldado-Magraner J, Smith MA, Yu BM (2024): MiSO: Optimizing brain stimulation to create neural population activity states. *Advances in Neural Information Processing Systems 37*. presented at the NeurIPS 2024.

31. Boutet A, Madhavan R, Elias GJB, Joel SE, Gramer R, Ranjan M, *et al.* (2021): Predicting optimal deep brain stimulation parameters for Parkinson's disease using functional MRI and machine learning. *Nat Commun* 12: 3043.

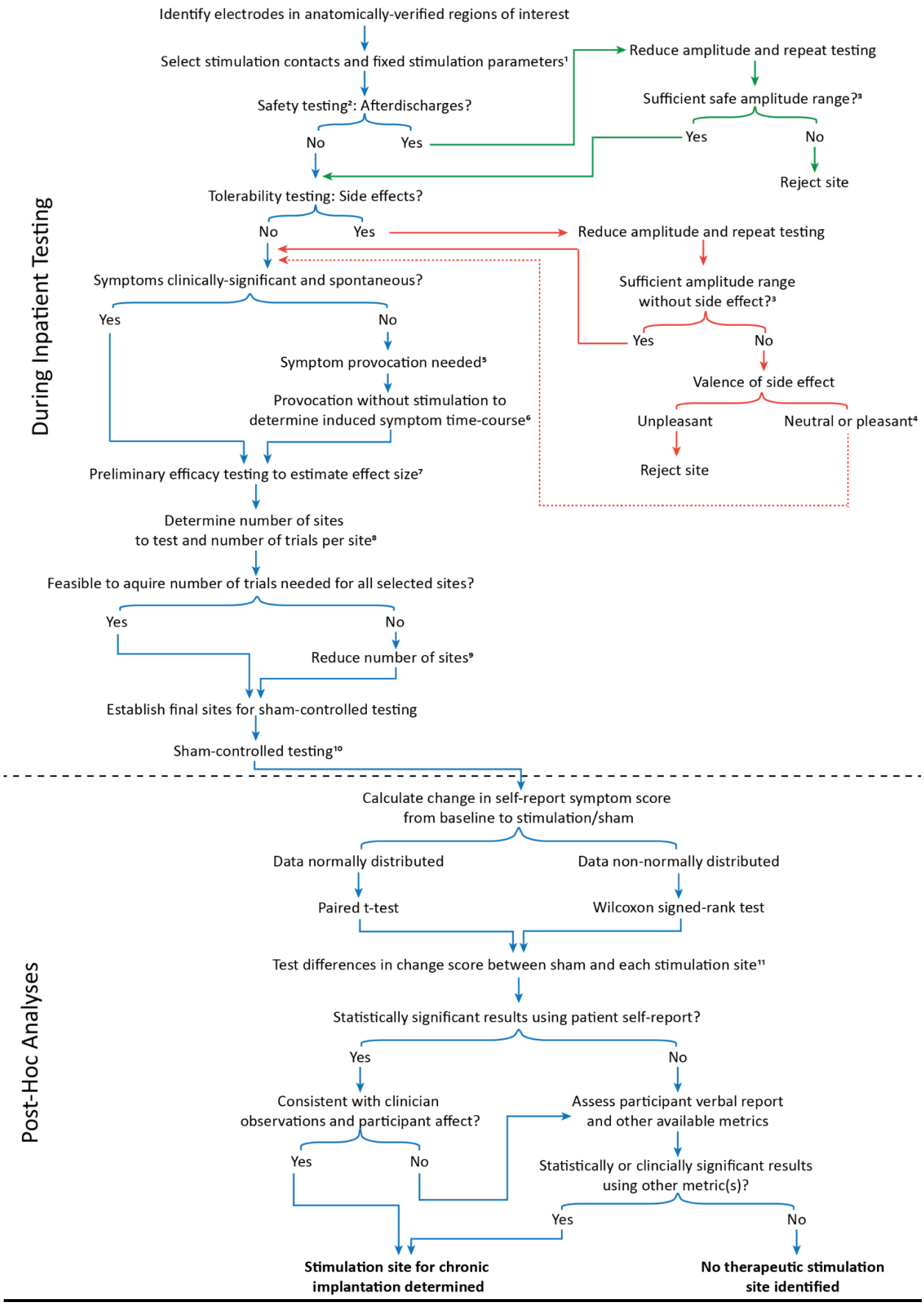

During Inpatient Testing
Identify electrodes in anatomically-verified regions of interest
Select stimulation contacts and fixed stimulation parameters[1]
Safety testing[2]: Afterdischarges?
No
Yes
Reduce amplitude and repeat testing
Sufficient safe amplitude range?[3]
Yes
No
Reject site
Tolerability testing: Side effects?
No
Yes
Reduce amplitude and repeat testing
Sufficient amplitude range without side effect?[3]
Yes
No
Valence of side effect
Unpleasant
Neutral or pleasant[4]
Reject site
Symptoms clinically-significant and spontaneous?
Yes
No
Symptom provocation needed[5]
Provocation without stimulation to determine induced symptom time-course[6]
Preliminary efficacy testing to estimate effect size[7]
Determine number of sites to test and number of trials per site[8]
Feasible to aquire number of trials needed for all selected sites?
Yes
No
Reduce number of sites[9]
Establish final sites for sham-controlled testing
Sham-controlled testing[10]
Post-Hoc Analyses
Calculate change in self-report symptom score from baseline to stimulation/sham
Data normally distributed
Data non-normally distributed
Paired t-test
Wilcoxon signed-rank test
Test differences in change score between sham and each stimulation site[11]
Statistically significant results using patient self-report?
Yes
No
Consistent with clinician observations and participant affect?
Yes
No
Assess participant verbal report and other available metrics
Statistically or clincially significant results using other metric(s)?
Yes
No
Stimulation site for chronic implantation determined
No therapeutic stimulation site identified

Figure 1. The Stimulation Testing Decision Tree provides a systematic and statistical approach for identifying therapeutic stimulation sites.

1: The parameter space is too large to exhaustively explore during inpatient testing. Select contact pairs for bipolar stimulation (anode and cathode) within each region of interest (ROI) based on co-registration of MRI and CT images. If multiple electrode contacts are within an ROI, different contact pairs can be tested for safety and tolerability. Also determine which parameters are to be kept fixed across regions; we recommend at least initially holding frequency and pulse-width constant across all sites tested.

2: Initial safety testing recommendation: 1s and then 3s at 1, 3, and 6mA using selected fixed stimulation parameters, per electrode contact pair (these parameters may need to be adjusted if using electrodes with small contact sizes or non-standard pitch).

3: Sufficient amplitude range may differ between cortical and subcortical targets. For both safety and side effects, we recommend rejecting sites if stimulation below 1mA for cortical and 0.8mA for subcortical sites is necessary. These amplitude ranges may need to be adjusted if using electrodes with small contact sizes or non-standard pitch.

4: The presence of side effects with stimulation may compromise participant blinding and subsequent double-blind randomized controlled trials.

5: Symptom provocation may include discussion of triggering life events (MDD), exposure to cues (OCD), or physical movement (CP). The specific activity should be tailored to a given participant's symptom profile and presentation.

6: Symptoms elicited by provocation may dissipate; the time-course of evoked symptoms should be measured and then monitored to inform when additional provocation instances are required.

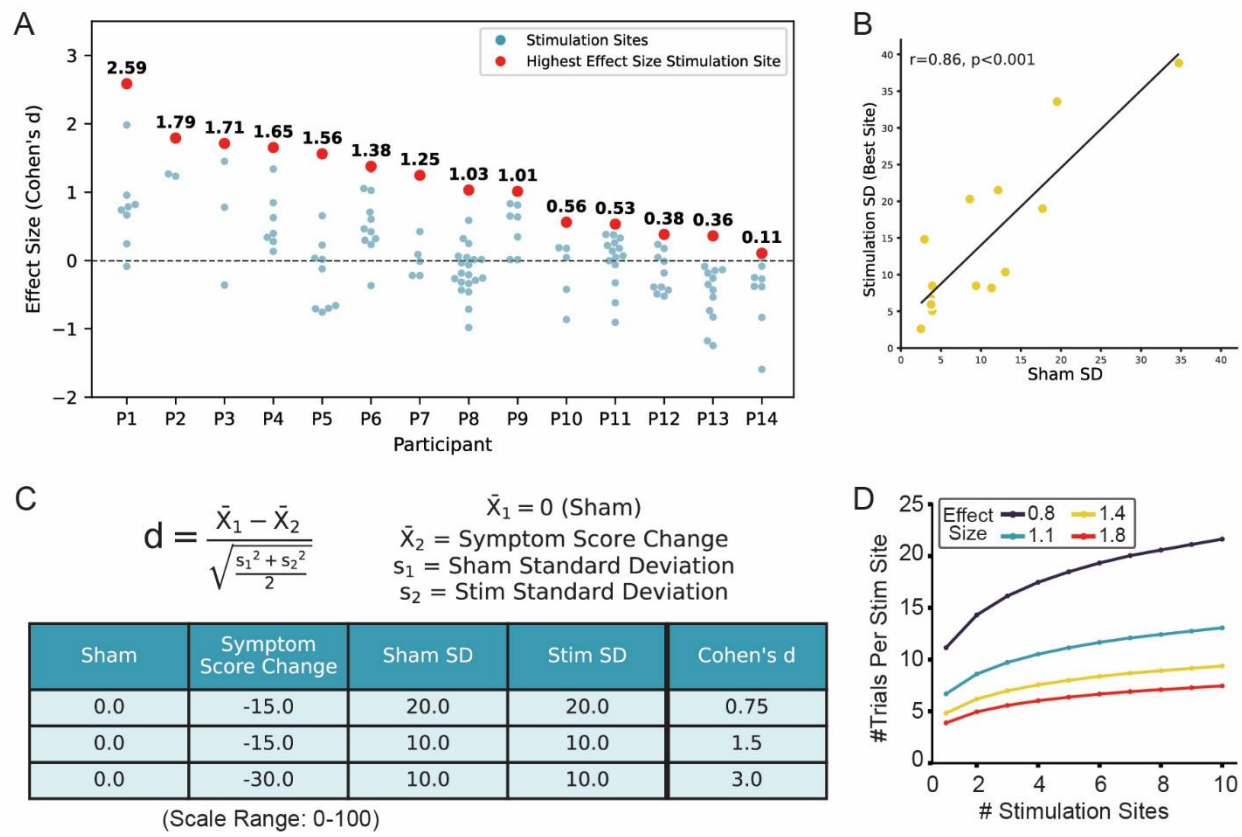
7: A subset of stimulation sites and parameters should be selected from those which survived safety and tolerability testing for preliminary efficacy testing. Estimated effect size is based on both the sham trial standard deviation as well as a change score value that indicates a clinical level of improvement.

8: The number of sites which can be tested should be determined by using Fig 2, which provides sample size based on estimated effect size and number of sites.

9: If the number of trials required for the sites that remain after safety and tolerability testing is not feasible for testing, the number of sites to be tested must be reduced. See main text for examples of methods to reduce the number of test sites.

10: Testing should be conducted using blocks within the same session consisting of at least one sham trial and up to one trial at each stimulation site. The sham trial is considered to be paired with each of the stimulation sites tested in that block. If multiple trials of sham are tested within one block, the scores should be averaged. See Supplementary Table 1.

11: Correct for multiple comparisons with, for example, Bonferroni's method.



Sham	Symptom Score Change	Sham SD	Stim SD	Cohen's d
0.0	-15.0	20.0	20.0	0.75
0.0	-15.0	10.0	10.0	1.5
0.0	-30.0	10.0	10.0	3.0

Figure 2. Effect Size Across Participants and Sample Size Estimation for Statistical Power

A: Cohen's d (verum stimulation vs. sham) for all stimulation sites tested in P1-P14. Red points indicate each participant's largest positive effect size, and blue points indicate effect sizes for all other stimulation sites tested. The dashed line at d = 0 represents no difference from sham. Cohen's d > 0 indicates improved symptoms relative to sham, whereas d < 0 indicates symptom worsening.

B: Strong positive correlation between each participant's sham SD and the SD at their best stimulation site ($r = 0.86$, $p < 0.001$ The regression line (black) shows a near one-to-one relationship (slope = 1.05, $p < 0.001$) with an intercept of 3.47.

C: Cohen's d formula and corresponding variables. The table provides three example calculations.

D: Number of trials required for statistical significance as a function of the number of brain regions, based on a paired t-test. Each line corresponds to a different effect size. Larger effect sizes require fewer trials to reach significance, while smaller effect sizes require more trials.