1 Sweetspot mapping in deep brain stimulation: 2 Strengths and limitations of current approaches

3

4 Running Title: DBS sweetspots: Current approaches. (32/40)

5 Till A. Dembek, MD¹, Juan Carlos Baldermann, MD¹, Jan-Niklas Petry-Schmelzer, MD¹,
6 Hannah Jergas¹, Harald Treuer, PhD², Veerle Visser-Vandewalle, MD², Haidar S. Dafsari, MD¹,
7 Michael T. Barbe, MD¹

8 ¹Department of Neurology, Faculty of Medicine, University of Cologne, Cologne, Germany

9 ²Department of Stereotactic and Functional Neurosurgery, Faculty of Medicine, University of
10 Cologne, Cologne, Germany

11 Funding: This study received no external funding.

12 Authorship Statement: TAD, JCB, JNPS, HJ & MTB developed the research design. TAD
13 was responsible for data acquisition, analysis and interpretation
14 of data. TAD drafted the manuscript. JCB, JNPS, HJ, HT, VVV,
15 HSD & MTB critically revised the manuscript for important
16 intellectual content. All authors approved the submitted
17 version of the manuscript.

18 Conflict of Interests: There were no conflicts of interest regarding this manuscript.
19 Full disclosures are included for each author in the respective
20 form.

21 Corresponding Author: Till A. Dembek, Department of Neurology, University Hospital of
22 Cologne, Kerpener Strasse 62, D-50937 Cologne, Germany.
23 Phone: +49 221 478 97602; Fax: +49 221 478 32139
24 eMail: till.dembek@uk-koeln.de

25 Word Counts:

26 Abstract: 216/300
27 Manuscript: 3968/4000
28 Introduction: 402
29 Methods: 1612
30 Results: 749
31 Discussion: 1201

32

33 Abstract

34 Objective: Open questions remain regarding the optimal target, or sweetspot, for deep brain
35 stimulation (DBS) in e.g. Parkinson's Disease. Previous studies introduced different methods
36 of mapping DBS effects to determine sweetspots. While having a direct impact on surgical
37 targeting and postoperative programming in DBS, these methods so far have not been
38 investigated in ground-truth data.

39 Materials & Methods: This study investigated five previously published DBS mapping
40 methods regarding their potential to correctly identify a ground-truth sweetspot. Methods
41 were investigated *in silico* in eight different use-case scenarios, which incorporated different
42 types of clinical data, noise, and differences in underlying neuroanatomy. Dice-coefficients
43 were calculated to determine the overlap between identified sweetspots and the
44 ground-truth. Additionally, out-of-sample predictive capabilities were assessed using the
45 amount of explained variance R^2 .

46 Results: The five investigated methods resulted in highly variable sweetspots. Methods
47 based on voxel-wise statistics against average outcomes showed the best performance
48 overall. While predictive capabilities were high, even in the best of cases Dice-coefficients
49 remained limited to values around 0.5, highlighting the overall limitations of sweetspot
50 identification.

51 Conclusions: This study highlights the strengths and limitations of current approaches to DBS
52 sweetspot mapping. Those limitations need to be taken into account when considering the
53 clinical implications. All future approaches should be investigated *in silico* before being
54 applied to clinical data.

55 Key words: Deep Brain Stimulation; Probabilistic Mapping; Sweetspot; Voxel-wise statistics;

56 Introduction

57 Deep brain stimulation (DBS) is an established treatment option for a variety of neurological
58 disorders. For example, electric stimulation of the subthalamic nucleus (STN) and its
59 surroundings via implanted electrodes can lead to a marked and lasting improvement of
60 motor¹ and non-motor symptoms² in Parkinson's disease (PD). However, DBS still is an
61 evolving therapy with new indications, new surgical targets, and new stimulation strategies
62 and techniques emerging every year³⁻⁵. To achieve the greatest benefit in each individual
63 patient, electrodes need to be implanted accurately in the optimal position and stimulation
64 parameters have to be finely adjusted to stimulate the exact, most favorable region.
65 Unfortunately, many open questions still remain regarding where these optimal stimulation
66 locations/regions are situated and whether they are comprised of particular
67 neuroanatomical regions i.e. nuclei or fiber tracts⁶⁻⁸. Determining optimal stimulation
68 regions thus is a crucial step to move the field of DBS forward and a variety of research
69 groups, employing heterogeneous strategies have made efforts in finding these 'sweetspots'
70 for different indications for almost a decade^{6,8-16}. In most cases, sweetspots were
71 determined by incorporating different data, consisting of 1) the locations of implanted
72 electrodes as determined from pre- and postoperative neuroimaging, 2) stimulation
73 parameters and models of electrical stimulation, and 3) clinical outcome parameters. By
74 then mapping clinical outcomes to the locations of stimulation over groups of patients and
75 employing different statistical methods or thresholding a sweetspot linked to most
76 advantageous outcome can be determined. Crucially, a variety of such mapping strategies
77 have been proposed but since the 'real' sweetspots are unknown, it is difficult to investigate
78 and compare their validity. Additionally, these methods sweetspots have neither been
79 validated in ground truth data nor in prospective clinical trials where these sweetspots are

80 explicitly targeted. While the results of DBS sweetspot-analyses have reached a wide
81 audience and may directly impact current concepts regarding the neuroanatomical origins of
82 DBS effects, surgical DBS targeting, and postoperative DBS programming, it remains unclear
83 whether the methods on which these sweetspots were based are able to produce reliable
84 results. The aim of this study is to address these dilemmas by comparing different mapping
85 strategies *in silico*, using a ‘ground-truth’ sweetspot. In doing so, we aim to give a
86 comprehensive overview about the strengths and limitations in existing strategies for
87 sweetspot-mapping in DBS as well as to provide guidelines to achieve more accurate and
88 comparable mapping results for future DBS trials.

89 Methods

90 Ethics & informed consent

91 This was an *in silico* trial which used no clinical data except retrospectively gathered, fully
92 anonymized locations of DBS leads. Consecutively no ethics vote or informed consent was
93 needed for this study.

94 Dataset

95 The dataset was based on electrode locations from n = 100 PD patients who had been
96 bilaterally implanted with DBS electrodes targeting the STN in our center. Electrode
97 locations had been determined¹⁷ using the Lead-DBS toolbox.¹⁸ Preoperative MRI and
98 postoperative CT images were coregistered linearly using Advanced Normalization Tools
99 (ANTs).¹⁹ Images were then normalized into the MNI space (2009b, nonlinear, asymmetric)
100 using the symmetric diffeomorphic registration approach (SyN) implemented in ANTs and
101 the ‘subcortical refine’ setting as implemented in Lead-DBS.^{19–21} Leads were identified using
102 the PaCER algorithm.²² The Medtronic 3389 lead model was chosen for all patients. For

further analysis, all leads were transferred to the right hemisphere. All steps were visually inspected by an experienced user (TAD).

Stimulation volumes

For each lead we calculated binarized stimulation volumes (or volumes of tissue activated, VTA). VTAs were estimated using the FastField algorithm as implemented in Lead-DBS using an electric-field threshold of 0.2 V/mm and an isotropic conductivity of 0.1 S/m^{23,24}. VTAs were calculated for each electrode on each lead with amplitudes ranging from 1 to 5 mA in 1 mA steps and a pulsewidth of 60 μ s. This resulted in 5 VTAs per electrode, 20 VTAs per lead and 4000 VTAs over all 100 patients. All VTAs were transferred to a 100x100x100 isotropic voxel-space with a voxel-size of 0.2x0.2x0.2 mm which covered the whole STN area.

‘Ground-truth’ sweetspot and clinical improvement

For this *in silico* study the sensorimotor subpart of the STN, as defined in the DISTAL atlas,²⁵ was defined as the ‘ground-truth’ sweetspot. We presumed that a higher stimulation overlap with the sensorimotor STN would be the sole predictor of ‘clinical’ outcome. A VTA which showed no overlap with the sensorimotor STN would thus lead to an improvement of 0 % while a VTA covering the whole sensorimotor STN would lead to an improvement of 100 %. In reality however, clinical outcome is of course influenced by a variety of other factors and often only depicted on low-resolution clinical scales. To address this issue, we also included an analysis where we added Gaussian noise to the calculated improvements based on their variance. Additionally, we conducted a separate analysis where we added the neighboring internal capsule (IC), as defined in the DISTAL atlas, as a structure associated with ‘side-effects’. We presumed that VTAs covering at least an arbitrary 16 mm³ or 2000 voxels of the internal capsule would cause a side-effect and thus could not be used. As a

result, this shifted the overall center of stimulation away from the neighboring internal capsule.

Stimulation Scenarios

There are two typical scenarios occurring in previous studies that tried to determine DBS sweetspots. Many studies were based on the clinical outcome of clinically-used, chronic stimulation settings. These settings, derived over multiple clinical visits, should, at least in theory, lead to the individually best outcome in each patient. In such studies, there is typically only one stimulation setting (two hemispheres) which can be used for sweetspot mapping and thus larger patient cohorts are needed^{9,12,13,15}. The second type of sweetspot mapping studies rely on stimulation data from acutely investigated stimulation settings, most often as part of so-called monopolar reviews. In these studies, multiple stimulation configurations per patient, often from each available electrode with increasing amplitudes are used. Such datasets naturally incorporate both; stimulation data with favorable, but also less desirable clinical outcome. Since a multitude of datapoints exist for each patient but stimulation testing can be burdensome, these studies often rely on smaller cohorts^{6,10,11,14}. For this study, we tried to reflect both scenarios in two different analyses.

For the first scenario ('Clinical Scenario') we selected only one of the calculated 20 VTAs per lead. In leads where no VTA caused an improvement of at least 50 %, the VTA associated with the highest improvement was chosen. In leads in which multiple VTAs were associated with an improvement greater than 50 %, the VTA with the highest ratio of improvement divided by stimulation amplitude was chosen. , thus reflecting clinical programming where in general the stimulation setting with the lowest amplitude is chosen when multiple settings lead to a satisfactory improvement. Having selected one VTA for each of the 200 leads, we

then randomly selected 100 leads/VTAs to reflect a typical cohort size for DBS mapping trials with long-term outcomes^{7,9,12,13,15}.

For the second scenario ('Monopolar Review Scenario') we used all of the 20 VTAs per lead, reflecting data generated during a monopolar review. However we limited our analysis to 20 randomly selected leads, resulting in 400 VTAs, which again reflects a cohort size typical of mapping trials based on monopolar review data^{6,11,14}.

Each of the two stimulation scenarios was investigated with only the sensorimotor STN or the sensorimotor STN + IC, each with and without Gaussian noise applied to the clinical outcomes resulting in eight scenarios in total (Figure 1).

Sweetspot mapping

For each of these eight scenarios we investigated $n = 5$ sweetspot mapping methods which have been developed and evolved over the course of almost a decade. An overview of these methods can be found in Table 1 and a detailed description of each method is included as Supplementary Material 1. In chronological order, 1) Butson et al. first mapped a sweetspot in PD patients using a VTA-based mean-effect image which was thresholded to select voxels with an average clinical improvement $> 50\%$ to define a sweetspot.¹⁰ 2) Cheung et al. defined a sweetspot for dystonia by investigating which voxels were most often covered by VTAs in patients with good DBS outcome.⁹ From their cohort, they only used the VTAs of the 75 % of patients with the best outcome and then selected all voxels which were stimulated by at least 75 % of these VTAs to define their sweetspot. In 2014, 3) Eisenstein et al. introduced voxel-wise statistics to the field of DBS mapping to define voxels associated with a significant improvement for different motor- and non-motor PD symptoms.¹³ Interestingly, their method was not based on VTAs but outcome was weighted for each voxel dependent on its distance to the active electrode. Then the resulting weighted outcomes for each voxel

were tested against zero to determine voxels which lead to a significant improvement. They also addressed the problem of multiple comparisons inherent to voxel-wise statistics by expanding their method with a post-hoc nonparametric permutation statistic. 4) Reich et al. used a VTA based mean-effect image and again used voxel-wise statistics to identify voxels which had a significantly higher outcome than the average outcome of all VTAs which did not stimulate that particular voxel.¹⁵ Clusters of significant voxels were identified using a fixed threshold of $n = 500$ voxels for cluster size with smaller clusters being discarded. The same year, 5) our group (Dembek et al.) used a method in monopolar review data of PD patients.⁶ A mean-effect image was generated and voxel-wise statistics were used to compare the outcomes in each voxel against average, amplitude-corrected outcomes of the cohort as a whole. Based on Eisenstein et al. we then used a nonparametric permutation statistic to correct for multiple comparisons and identify valid clusters of significant voxels. For this study, these five mentioned methods were recreated in Matlab® (The Mathworks, Natick, USA). Other methods presented in the past were excluded from our analysis for being highly similar to either the method by Eisenstein et al.¹², Dembek et al.¹¹ or for using clinical data not comparable with our stimulation scenarios¹⁴. For all analyses, only voxels which were at least covered by $n = 16$ VTAs were regarded, to allow reliable statistics^{6,11}.

Statistical analysis and outcomes

Each of the eight scenarios was repeated 30 times, each time using randomly selected DBS leads. For each of the 30 iterations we analyzed the following outcome parameters: To address the overall spatial concordance between the sensorimotor STN and the sweetspots detected by each method, we calculated the Sørensen-Dice-coefficient. The Sørensen-Dice-coefficient (Dice-coefficient for short) is determined by calculating two times the size of the overlap between two datasets and dividing this by the sum of the individual

sizes of the two datasets. It thus ranges from Dice = 0, meaning no overlap between datasets, to Dice = 1, indicating perfectly congruent datasets. To assess the clinical relevance of the sweetspots, we also determined their predictive capabilities in out of sample data. Prediction was assessed by calculating the overlap between the sweetspot determined during one iteration with the VTAs from another randomly selected iteration. Then linear regression analysis was performed to determine what amount of variance (R^2) in clinical outcome could be explained by the overlap with the sweetspot. Dice-coefficients and R^2 -values were compared between the five methods using a Friedman test with post-hoc pairwise comparisons using the Wilcoxon signed-rank test, Bonferroni-corrected for the number of comparisons ($n = 10$). Additionally, the sizes of sweetspots were calculated and compared to the size of the sensorimotor STN for each mapping method using Wilcoxon signed-rank tests, again Bonferroni-corrected for the number of comparisons ($n = 5$). All results are shown as medians $\pm$ range with Bonferroni-corrected p-values which were deemed significant if < 0.05 .

Further analyses

To describe the internal consistency of sweetspots generated by each method, we calculated Dice-coefficients between all sweetspots generated by one method during the 30 iterations for each scenario. Additionally, we calculated 'average sweetspots' for each method and each scenario, by determining which voxels were part of the respective sweetspot in at least 50 % of the 30 iterations.

Results

Clinical Scenario

Results for the Clinical Scenario are summarized in Figure 2 (STN) and Figure 3 (STN + IC). The Friedman test revealed significant differences between methods for all analyses. Importantly, the post-hoc nonparametric permutation analysis of the Eisenstein et al. method discarded the detected sweetspots in scenarios where noise had been added. When comparing Dice-coefficients in our clinical scenario without added noise and without the internal capsule as an exclusion zone the Dembek et al. method significantly outperformed all other methods. When adding Gaussian noise to the clinical outcomes, the method by Cheung et al. had the highest Dice-coefficients. After excluding VTAs overlapping the internal capsule, the methods by Reich et al. and Dembek et al. both had significantly higher Dice-coefficients than other methods. The predictive capabilities, as measured by the amount of explained variance R^2 were highest for the methods by Reich et al. and Dembek et al., which were able to explain around 80 % of the variance in clinical outcome. The sweetspots determined by the Butson et al. and Eisenstein et al. methods were significantly larger than the sensorimotor STN, while the method by Cheung et al. consistently led to smaller sweetspots.

Monopolar Review Scenario

Results for the Monopolar Review Scenario are summarized in Figure 4 (STN) and Figure 5 (STN + IC). Surprisingly, the method presented by Cheung et al. was not able to identify sweetspots in our data since the maximum number of overlaps of high outcome VTAs never reached the predefined 75% threshold. Also, sweetspots detected by the Eisenstein et al. method were again discarded during its nonparametric permutation analysis in scenarios with added noise. Dice-coefficients and predictive capabilities as described by the amount of explained variance R^2 were significantly higher in the Dembek et al. method compared to all other remaining methods for all variants of the monopolar review scenario (with/without

internal capsule; with/without noise). About 90 % of variance in outcomes could be explained using these sweetspots. The Reich et al. method showed the second highest Dice-coefficients and predictive capabilities in all cases.

General sweetspot characteristics

Overall methods resulted in characteristic sweetspots which were often consistent within each method (see Figure 6 for average sweetspots and Table 2 for within-methods Dice-coefficients). The Butson et al. method, based on a 50 % improvement threshold applied to the Mean-image often generated ‘shell-like’ sweetspots which emphasized on the posterior border of the investigated area. Results were quite inconsistent as illustrated by the reduced Dice-coefficient within the method, especially when noise was added. The Cheung et al. method, based on a fixed threshold applied to the N-image of responder VTAs resulted in very consistent, sphere-like sweetspots, which were close to the center of the overall investigated area and mostly lay within the sensorimotor STN. When using the Eisenstein et al. method with its voxel-wise statistics against zero, sweetspots tended to be large, round volumes, covering large parts of the investigated area. While there was a high consistency within the method for different iterations, the sweetspots lacked confinement to the actual sensorimotor STN. Reich et al. and Dembek et al. methods, which use voxel-wise statistics against average outcomes, resulted in quite similar types of sweetspots. Those were mainly centered on the sensorimotor STN with some overspill outside its medial and lateral border. Consistency within these methods sometimes proved to be an issue when noise was added.

Effects of post-hoc analyses

Three of the investigated methods, namely Eisenstein et al., Reich et al., and Dembek et al., used post-hoc analyses to correct for multiple voxel-wise testing (see Table 3 for detailed

results). The Eisenstein et al. method, which evaluates the overall significance of the significant voxels against the overall significance of $n = 200$ permuted datasets discarded the whole sweetspot in all scenarios with added noise – and in none of the scenarios without added noise. The Reich et al. method, which discarded clusters of significant voxels whenever the size of the cluster was below $n = 500$ voxels mostly discarded scattered voxels and clusters consisting of fewer than 10 voxels. Such removal of scattered voxels mainly occurred in scenarios with added noise. The majority of those voxels was outside the ground-truth sweetspot. The Dembek et al. method, which ranked the overall significance of clusters of significant voxels against the most significant clusters of $n = 1000$ permuted datasets, also removed mostly small, scattered clusters. Cluster removal occurred more frequently than during the Reich et al. method, but again most removed voxels were outside the ground-truth sweetspot.

Discussion

This study provides insight into the question of how different methods of DBS mapping influence the resulting sweetspots. Using ground-truth data, this study explores the possibilities and especially the limitations for different approaches. As the main finding, different previously published approaches results in a variety of different DBS sweetspots when used in the same dataset and results can diverge widely even within methods for different use-case scenarios.

The investigated methods were developed over almost a decade, with newer methods heavily relying on their forebears. As a result of this evolution, we indeed found that the most recent approaches by Reich et al.¹⁵ and Dembek et al.⁶, which employed voxel-wise statistics comparing the outcomes of each voxel against an average of other outcomes in the dataset, provided the best results. Overall, these methods provided sweetspots with the

highest Dice-coefficients (i.e. similarity between computed and actual ground-truth sweetspot) and the highest predictive capabilities. A correction for stimulation amplitude, as implemented in Dembek et al.⁶, seemed to further improve results, especially with monopolar review data in which amplitude naturally plays a major role. Post-hoc analyses i.e. cluster thresholding or nonparametric permutation statistics helped to remove some scattered clusters of voxels but did not affect the results in this study in a major way. In real-life data however, false positives may play a larger role and post-hoc correction thus might be more important.

Employing a voxel-wise statistical analysis against zero, as implemented in the Eisenstein et al.¹³ method, consistently led to large parts of the investigated area being highlighted as significant. This is not surprising due to the nature of our data which only included clinical improvements and not clinical worsening. However, worsening of e.g. motor symptoms due to DBS is also rarely observed in clinical reality, so assuming zero improvement as the null-hypothesis for voxel-wise statistical analysis indeed might be too liberal and will provide less meaningful clinical insights. This reasoning is supported by the low amount of explained variance in our prediction analysis and the fact that the post-hoc analysis implemented in the method itself removed a lot of the resulting sweetspots as possible false-positives¹¹.

The approach by Cheung et al.⁹, which investigated where most responder VTAs overlap, provided relatively consistent results in clinical data which remained comparably unaffected by the addition of noise. Of note, this method underestimated the size of the sweetspot. Moreover, since this approach tended severely towards the overall center of stimulation, it performed worse when the center of stimulation was moved away from the sweetspot, e.g. in the scenarios with the added internal capsule. Additionally, this approach did not work

314 with monopolar review data – at least not when using the predefined overlap threshold of
315 75 %.

316 Importantly, the maximum achievable Dice-coefficients even in the best cases were around
317 0.5, meaning that there was still a significant mismatch between the calculated sweetspots
318 and the anatomical ground-truth. This result highlights the general limitations of current
319 approaches to DBS sweetspot mapping regarding neuroanatomical accuracy. These
320 limitations become even more important when one considers that the many other sources
321 of noise present in clinical reality, like e.g. imaging-inaccuracies, uncertainties regarding the
322 VTA-model, or the largely unknown underlying electric tissue properties, did not factor into
323 this *in silico* analysis. Additionally, one has to keep in mind that as a general limitation only
324 those parts of a sweetspot can be identified which are covered by enough stimulation data.
325 In reality, it might of course happen that the ‘real sweetspot’ is so far off target, that it
326 cannot be discovered at all using sweetspot mapping.

327 On the other hand, the Dembek et al. and Reich et al. methods could explain large amounts
328 of variance during the out-of-sample prediction analysis – highlighting their potential use in
329 future applications like computer-guided DBS programming. Taken together, the mismatch
330 between suboptimal anatomical accuracy and high predictive capabilities also sheds light
331 onto another important fact: namely that even if a sweetspot is able to predict clinical
332 outcome, it does not mean that its anatomical location is necessarily linked to the underlying
333 neuroanatomical structure causing the clinical effect. This is of particular importance when
334 drawing conclusions about the underlying neuroanatomy or networks from such
335 analyses^{7,8,11,26–28}.

336 Towards an improved DBS mapping strategy

337 When considering the results of this study as well as the many possible sources of noise in
338 DBS sweetspot mapping, it seems quite difficult to develop a strategy which can sufficiently
339 address all these factors. In our opinion, the ideal mapping strategy should not rely on fixed
340 and arbitrary thresholds. As demonstrated in this study, methods based on arbitrary
341 thresholds are highly sensitive to differences in use-case scenarios. Even if such thresholds
342 were defined in ground-truth models, one could never be sure that these models genuinely
343 reflect clinical reality. Voxel-wise statistics investigating outcome in a certain voxel against
344 (average) clinical outcomes in other DBS settings thus seem to be the most promising
345 approach at the moment. Such methods do not rely on thresholding and, as shown in our
346 analyses, seem to provide relatively consistent results for different scenarios. A possible
347 avenue for further improvement might be using more advanced statistical methods for
348 voxel-wise testing. Instead of t-tests^{12,13,15} or Wilcoxon signed-rank tests^{6,11}, voxel-wise
349 mixed-effect models might for example be a promising strategy to account for clinical
350 covariates like symptom severity or medication. They could also help to expand the relatively
351 narrow scope of the 'VTA' itself by incorporating other factors like stimulation frequency or
352 pulse-width, different stimulation configurations like anodal, bipolar or multi-electrode
353 stimulation^{29,30}, or axon properties like activation thresholds or directionality. Regarding
354 post-hoc corrections for multiple comparisons, we again think that threshold-based methods
355 like cluster thresholding should be avoided. While of course being highly inefficient from a
356 computational point of view, nonparametric permutation statistics have the important
357 advantage of being comparatively free of underlying assumptions. Another important
358 question is whether binary sweetspots are the optimal solution for describing areas with
359 good stimulation outcome. By binarizing DBS mapping results into a however well-defined
360 sweetspot, a lot of the information about spatial gradients/contrasts is lost. While binarized

sweetspots are of course useful and easily understandable approximations, it remains to be seen if more complex metrics might be better suited for e.g. outcome prediction analyses¹⁵.

No matter what methods will be developed and employed in the future, this study hopefully highlights the importance of ground-truth, *in silico* testing. Many previous DBS sweetspot studies have reached a broad audience, directly influencing our assumptions of where optimal stimulation outcome might be achieved. They thus have the potential to impact both: electrode placement and postoperative DBS programming. This highlights the importance of first examining a method's viability in ground-truth data before using this method to draw clinically relevant conclusions. The same also holds true for new approaches to DBS mapping which incorporate for example functional and/or structural connectomic data^{7,26,31}.

Conclusion

The abilities of current DBS mapping methods to reliably detect sweetspots for motor outcomes are limited and potentially influenced by a variety of factors. While methods have evolved over the years, there is still much room for improvement. *In silico* validation should be used before drawing clinical conclusions from new DBS mapping methods and researchers should be aware of the respective strengths and limitations of each method.

References

1. Xie C-L, Shao B, Chen J, Zhou Y, Lin S-Y, Wang W-W. Effects of neurostimulation for advanced Parkinson's disease patients on motor symptoms: A multiple-treatments meta-analysis of randomized controlled trials. *Sci Rep*. 2016;6:25285. doi:10.1038/srep25285
2. Jost ST, Sauerbier A, Visser-Vandewalle V, et al. A prospective, controlled study of non-motor effects of subthalamic stimulation in Parkinson's disease: results at the 36-month follow-up. *J Neurol Neurosurg Psychiatry*. 2020;91(7):687-694. doi:10.1136/jnnp-2019-322614
3. Harmsen IE, Elias GJB, Beyn ME, et al. Clinical trials for deep brain stimulation: Current state of affairs. *Brain Stimulation*. 2020;13(2):378-385. doi:10.1016/j.brs.2019.11.008
4. Ramirez-Zamora A, Giordano J, Gunduz A, et al. Proceedings of the Seventh Annual Deep Brain Stimulation Think Tank: Advances in Neurophysiology, Adaptive DBS, Virtual Reality, Neuroethics and Technology. *Front Hum Neurosci*. 2020;14. doi:10.3389/fnhum.2020.00054
5. Lozano AM, Lipsman N, Bergman H, et al. Deep brain stimulation: current challenges and future directions. *Nat Rev Neurol*. 2019;15(3):148-160. doi:10.1038/s41582-018-0128-2
6. Dembek TA, Roediger J, Horn A, et al. Probabilistic Sweetspots Predict Motor Outcome for DBS in Parkinson's Disease. *Ann Neurol*. Published online August 3, 2019. doi:10.1002/ana.25567
7. Horn A, Reich M, Vorwerk J, et al. Connectivity Predicts deep brain stimulation outcome in Parkinson disease. *Ann Neurol*. 2017;82(1):67-78. doi:10.1002/ana.24974
8. Akram H, Sotiropoulos SN, Jbabdi S, et al. Subthalamic deep brain stimulation sweet spots and hyperdirect cortical connectivity in Parkinson's disease. *NeuroImage*. 2017;158:332-345. doi:10.1016/j.neuroimage.2017.07.012
9. Cheung T, Noecker AM, Alterman RL, McIntyre CC, Tagliati M. Defining a therapeutic target for pallidal deep brain stimulation for dystonia. *Ann Neurol*. 2014;76(1):22-30. doi:10.1002/ana.24187
10. Butson CR, Cooper SE, Henderson JM, Wolgamuth B, McIntyre CC. Probabilistic analysis of activation volumes generated during deep brain stimulation. *NeuroImage*. 2011;54(3):2096-2104. doi:10.1016/j.neuroimage.2010.10.059
11. Dembek TA, Barbe MT, Åström M, et al. Probabilistic mapping of deep brain stimulation effects in essential tremor. *NeuroImage: Clinical*. 2017;13:164-173. doi:10.1016/j.nicl.2016.11.019
12. Gourisankar A, Eisenstein SA, Trapp NT, et al. Mapping movement, mood, motivation and mentation in the subthalamic nucleus. *Royal Society Open Science*. 2018;5(7):171177. doi:10.1098/rsos.171177

13. Eisenstein SA, Koller JM, Black KD, et al. Functional anatomy of subthalamic nucleus stimulation in Parkinson disease. *Ann Neurol*. 2014;76(2):279-295. doi:10.1002/ana.24204
14. Nguyen TAK, Nowacki A, Debove I, et al. Directional stimulation of subthalamic nucleus sweet spot predicts clinical efficacy: Proof of concept. *Brain Stimul*. 2019;12(5):1127-1134. doi:10.1016/j.brs.2019.05.001
15. Reich MM, Horn A, Lange F, et al. Probabilistic mapping of the antidystonic effect of pallidal neurostimulation: a multicentre imaging study. *Brain*. Published online March 8, 2019. doi:10.1093/brain/awz046
16. Phibbs FT, Pallavaram S, Tolleson C, D'Haese P-F, Dawant BM. Use of efficacy probability maps for the post-operative programming of deep brain stimulation in essential tremor. *Parkinsonism & Related Disorders*. 2014;20(12):1341-1344. doi:10.1016/j.parkreldis.2014.09.004
17. Dembek TA, Hoevels M, Hellerbach A, et al. Directional DBS leads show large deviations from their intended implantation orientation. *Parkinsonism Relat Disord*. 2019;67:117-121. doi:10.1016/j.parkreldis.2019.08.017
18. Horn A, Li N, Dembek TA, et al. Lead-DBS v2: Towards a comprehensive pipeline for deep brain stimulation imaging. *NeuroImage*. Published online September 1, 2018. doi:10.1016/j.neuroimage.2018.08.068
19. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain. *Med Image Anal*. 2008;12(1):26-41. doi:10.1016/j.media.2007.06.004
20. Schönecker T, Kupsch A, Kühn AA, Schneider G-H, Hoffmann K-T. Automated optimization of subcortical cerebral MR imaging-atlas coregistration for improved postoperative electrode localization in deep brain stimulation. *AJNR Am J Neuroradiol*. 2009;30(10):1914-1921. doi:10.3174/ajnr.A1741
21. Fonov V, Evans AC, Botteron K, Almli CR, McKinsty RC, Collins DL. Unbiased Average Age-Appropriate Atlases for Pediatric Studies. *Neuroimage*. 2011;54(1):313-327. doi:10.1016/j.neuroimage.2010.07.033
22. Husch A, V. Petersen M, Gemmar P, Goncalves J, Hertel F. PaCER - A fully automated method for electrode trajectory and contact reconstruction in deep brain stimulation. *Neuroimage Clin*. 2017;17:80-89. doi:10.1016/j.nicl.2017.10.004
23. Baniasadi M, Proverbio D, Gonçalves J, Hertel F, Husch A. FastField: An Open-Source Toolbox for Efficient Approximation of Deep Brain Stimulation Electric Fields. *bioRxiv*. Published online March 5, 2020:2020.03.03.974642. doi:10.1101/2020.03.03.974642
24. Astrom M, Diczfalusy E, Martens H, Wardell K. Relationship between Neural Activation and Electric Field Distribution during Deep Brain Stimulation. *IEEE Trans Biomed Eng*. 2015;62(2):664-672. doi:10.1109/TBME.2014.2363494

25. Ewert S, Plettig P, Li N, et al. Toward defining deep brain stimulation targets in MNI space: A subcortical atlas based on multimodal MRI, histology and structural connectivity. *NeuroImage*. Published online May 20, 2017. doi:10.1016/j.neuroimage.2017.05.015
26. Li N, Baldermann JC, Kibleur A, et al. A unified connectomic target for deep brain stimulation in obsessive-compulsive disorder. *Nature Communications*. 2020;11(1):3364. doi:10.1038/s41467-020-16734-3
27. Petry-Schmelzer JN, Krause M, Dembek TA, et al. Non-motor outcomes depend on location of neurostimulation in Parkinson's disease. *Brain*. 2019;142(11):3592-3604. doi:10.1093/brain/awz285
28. Irmen F, Horn A, Mosley P, et al. Left Prefrontal Connectivity Links Subthalamic Stimulation with Depressive Symptoms. *Ann Neurol*. 2020;87(6):962-975. doi:10.1002/ana.25734
29. Anderson DN, Duffley G, Vorwerk J, Dorval A (Chuck), Butson CR. Anodic Stimulation Misunderstood: Preferential Activation of Fiber Orientations with Anodic Waveforms in Deep Brain Stimulation. *J Neural Eng*. Published online 2018. doi:10.1088/1741-2552/aae590
30. Slopsema JP, Peña E, Patriat R, et al. Clinical deep brain stimulation strategies for orientation-selective pathway activation. *J Neural Eng*. 2018;15(5):056029. doi:10.1088/1741-2552/aad978
31. Baldermann JC, Melzer C, Zapf A, et al. Connectivity Profile Predictive of Effective Deep Brain Stimulation in Obsessive-Compulsive Disorder. *Biol Psychiatry*. 2019;85(9):735-743. doi:10.1016/j.biopsych.2018.12.019

Table 1 – Overview of sweetspot mapping methods

Method	Based on	Mapped Outcome	Voxel-wise Statistics	0-Hypothesis for voxel-wise statistics	Amplitude correction	Sweetspot threshold	Correction for multiple Comparisons
Butson et al. 2011	VTAs	Mean Outcome per Voxel	No	-	No	Mean improvement per Voxel > 50 %	No
Cheung et al. 2014	VTAs	No. of Overlaps per Voxel	No	-	No	At least 75% of responder VTAs overlap	No
Eisenstein et al. 2014	Coordinates	Mean Outcome, weighted by distance to electrode	One sample T-test	Zero	No	Significance in Voxel-wise statistics	Nonparametric permutation statistic (whole image)
Reich et al. 2019	VTAs	Mean Outcome per Voxel	Two sample T-Test	Average outcome of unconnected VTAs	No	Significance in Voxel-wise statistics	Cluster size > 500 voxels
Dembek et al. 2019	VTAs	Mean Outcome per Voxel	Wilcoxon signed-rank test	Average outcome	Yes	Significance in Voxel-wise statistics	Nonparametric permutation statistic (cluster-based)

Table 2 – Dice-coefficients within methods

			Butson et al.	Cheung et al.	Eisenstein et al.	Reich et al.	Dembek et al.
Clinical Scenario	STN	no noise	0.89 (0.80-0.95)	0.92 (0.79-0.96)	0.96 (0.91-0.97)	0.82 (0.50-0.91)	0.83 (0.60-0.92)
		with noise	0.63 (0.07-0.88)	0.91 (0.77-0.96)	-	0.34 (0.00-0.83)	0.48 (0.04-0.79)
	STN + IC	no noise	0.90 (0.80-0.96)	0.88 (0.77-0.95)	0.95 (0.92-0.98)	0.88 (0.73-0.94)	0.88 (0.77-0.93)
		with noise	0.33 (0.02-0.86)	0.84 (0.66-0.94)	-	0.74 (0.51-0.88)	0.65 (0.39-0.86)
Monopolar Review Scenario	STN	no noise	0.57 (0.29-0.90)	-	0.87 (0.75-0.92)	0.82 (0.65-0.92)	0.81 (0.61-0.92)
		with noise	0.33 (0.06-0.77)	-	-	0.81 (0.58-0.91)	0.79 (0.55-0.91)
	STN + IC	no noise	0.57 (0.29-0.83)	-	0.86 (0.70-0.93)	0.79 (0.56-0.92)	0.75 (0.42-0.90)
		with noise	0.27 (0.04-0.68)	-	-	0.77 (0.52-0.89)	0.67 (0.41-0.84)

Legend: Median (range) Dice-coefficients within each of the five investigated methods (columns) and each of the eight scenarios (rows) are shown.

Table 3 – Effects of post-hoc analyses

			Eisenstein et al.			Reich et al.			Dembek et al.		
			Percentage of iterations in which voxels were removed	Median (range) number of removed voxels	Removed voxels outside the ground truth sweetspot	Percentage of iterations in which voxels were removed	Median (range) number of removed voxels	Removed voxels outside the ground truth sweetspot	Percentage of iterations in which voxels were removed	Median (range) number of removed voxels	Removed voxels outside the ground truth sweetspot
Clinical Scenario	STN	no noise	0 %	-	-	17 %	1 (1-2)	100 %	60 %	3 (1-13)	93 %
		with noise	83 %	29211 (28011-30225)	78 %	80 %	3 (1-463)	25 %	93 %	12 (1-80)	84 %
	STN + IC	no noise	0 %	-	-	10 %	1 (1-2)	100 %	20 %	1 (1-2)	100 %
		with noise	100 %	27884 (25355-30067)	76 %	77 %	3 (1-11)	80 %	83 %	4 (1-13)	67 %
Monopolar Review Scenario	STN	no noise	0 %	-	-	10 %	1 (1-2)	100 %	3 %	1 (1-1)	100 %
		with noise	100 %	45957 (38402-53222)	85 %	27 %	2.5 (1-7)	100 %	37 %	3 (1-4)	100 %
	STN + IC	no noise	0 %	-	-	13 %	1.5 (1-4)	100 %	0 %	-	-
		with noise	100 %	37071 (27844-49257)	82 %	60 %	1.5 (1-19)	100 %	60 %	3 (1-13)	91 %

Legend: Effects of the respective post-hoc analyses incorporated in Eistenstein et al., Reich et al., and Dembek et al. (columns) for each of the eight scenarios (rows) are shown. Median values and the range of values are provided.

Figure 1 – Overview of the investigated scenarios

Legend: Schematic overview of the investigated stimulation scenarios. The sensorimotor STN is shown in blue, the STN is shown as outline, and the internal capsule (IC) in yellow. Each of the eight scenarios was repeated 30 times, each time with randomly selected leads.

Figure 2 – Clinical Scenario; STN

Legend: Results of the clinical scenario without the internal capsule, both without (first row) and with noise added to the clinical outcomes (second row). Boxplots are shown for Dice-coefficients (first column), the predictive capabilities (second column) and the relative sizes of sweetspots with respect to the sensorimotor STN (third column). Boxplots depict the median, interquartile range and range. The statistical significance of pairwise comparisons is indicated using asterisks for Dice-coefficients and the amount of explained variance. For size, the statistical significance between each method and the size of the sensorimotor STN are indicated using asterisks. Asterisks indicate significance for p-values corrected for multiple comparisons: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Figure 3 – Clinical Scenario; STN + IC

Legend: Results of the clinical scenario with the internal capsule (IC), both without (first row) and with noise added to the clinical outcomes (second row). Boxplots are shown for Dice-coefficients (first column), the predictive capabilities (second column) and the relative sizes of sweetspots with respect to the sensorimotor STN (third column). Boxplots depict the median, interquartile range and range. The statistical significance of pairwise comparisons is indicated using asterisks for Dice-coefficients and the amount of explained variance. For size, the statistical significance between each method and the size of the sensorimotor STN are indicated using asterisks. Asterisks indicate significance for p-values corrected for multiple comparisons: * p < 0.05; ** p < 0.01; *** p < 0.001.

Figure 4 – Monopolar Review Scenario; STN

Legend: Results of the monopolar review scenario without the internal capsule, both without (first row) and with noise added to the clinical outcomes (second row). Boxplots are shown for Dice-coefficients (first column), the predictive capabilities (second column) and the relative sizes of sweetspots with respect to the sensorimotor STN (third column). Boxplots depict the median, interquartile range and range. The statistical significance of pairwise comparisons is indicated using asterisks for Dice-coefficients and the amount of explained variance. For size, the statistical significance between each method and the size of the sensorimotor STN are indicated using asterisks. Asterisks indicate significance for p-values corrected for multiple comparisons: * p < 0.05; ** p < 0.01; *** p < 0.001.

Figure 5 – Monopolar Review Scenario; STN + IC

Legend: Results of the monopolar review scenario with the internal capsule (IC), both without (first row) and with noise added to the clinical outcomes (second row). Boxplots are shown for Dice-coefficients (first column), the predictive capabilities (second column) and the relative sizes of sweetspots with respect to the sensorimotor STN (third column). Boxplots depict the median, interquartile range and range. The statistical significance of pairwise comparisons is indicated using asterisks for Dice-coefficients and the amount of explained variance. For size, the statistical significance between each method and the size of the sensorimotor STN are indicated using asterisks. Asterisks indicate significance for p-values corrected for multiple comparisons: * p < 0.05; ** p < 0.01; *** p < 0.001.

Figure 6 – Average Sweetspots

Legend: Average sweetspots (red) signify all voxels localized within the sweetspot at least 50 % of times for the respective method (columns) and scenario (rows). Grey volumes on the other hand encompass all voxels which lay inside a sweetspot at least once. The sensorimotor STN (blue) signifies the ground-truth sweetspot. The whole STN is shown as black outline.