Chapter 38

The Design of Clinical Studies for Neuromodulation

Wael F. Asaad, Peter M. Lauro, Shane Lee

[For Springerlink]

Abstract

Creating successful new neuromodulation therapies requires innovative trial design and the balancing of a wide variety of complex factors. The ultimate goal is to advance our ability to restore appropriate brain function through the targeted manipulation of neural circuits and, to this end, clinical trials should be just as informative in failure as they are in success. Some aspects of neuromodulation trials are common to all clinical studies, while other aspects are fairly specific to invasive neurosurgical interventions for complex neurological and psychiatric disease. This chapter examines the major factors to be considered in constructing clinical studies to investigate the various forms and applications of novel neuromodulation strategies.

Keywords:

Introduction

The more we understand about the electrophysiology and circuits underlying neuropsychiatric disease, the greater the possibility that focal neuromodulation will be a viable therapeutic strategy. Although the administration of drugs can certainly be a form of neuromodulation, we draw a distinction between chemically-targeted and anatomically-targeted forms of neural systems manipulation, and will use the term “neuromodulation” to refer here specifically to the latter.

Compared to the systemic delivery of drugs, anatomically-based neuromodulation strategies are not limited by the existing distributions of molecular targets. While there may be disease entities that are strictly defined by a small number of reversible molecular derangements, there are others — stroke and traumatic brain or spine injury being the most notable — where the pathology does not conform to existing cellular and molecular boundaries. Other diseases, such as epilepsy, may in some cases begin as a circumscribed molecular derangement, but additional circuits may be recruited over time in a manner that does not respect the boundaries of the inciting pathologic entity [1, 2]. In such cases, a treatment approach is required that addresses the resulting neural dysfunction in a manner that is tailored to the type and extent of circuit pathology and that is not limited by the naturally-occurring distributions of molecular targets. Parenthetically, while new molecular targets might be introduced in an anatomically-targeted manner for interaction with systemically-administered medications (e.g., designer receptors for designer drugs), such a strategy would be subject to many of the considerations discussed here.

Of course, current neuromodulation strategies often require invasive techniques and, like systemically administered medications, are likely to have limited specificity at the target as well as poorly defined, extended effects at adjacent, upstream, and downstream sites. Nevertheless,

the promise of spatially-targeted neuromodulation techniques is ever-increasing anatomical and functional specificity beyond that provided by nature's endowment of the brain with particular molecules in particular distributions.

The process of designing and testing new neuromodulation strategies shares some features with the development of drug-based therapies, but a focus on neuromodulation also introduces novel factors into the process that may be challenges and/or opportunities. Here, we will examine the common and distinctive considerations of clinical studies that seek to investigate potential neuromodulatory approaches to nervous system dysfunction. While the primary hope of all such investigations is to establish the potential or actual clinical utility of a given interventional strategy, the reality is that such studies are more likely ultimately to fail than to succeed in achieving the primary therapeutic endpoint. With this in mind, the design of a clinical investigation must be optimized not only to maximize the probability of success, but also to learn as much from failure as from success, so that future endeavors are able to build solidly upon new knowledge in a step-wise and ultimately fruitful manner.

Defining the Scope and Power of a Study

There have been relatively few sufficiently large-scale, prospective, double-blinded, controlled neuromodulation trials (Table 38.1). The expensive nature of neuromodulation therapies — whether due to the cost of implanted devices (e.g., deep brain stimulation systems), therapeutic delivery systems (e.g., focused ultrasound), adjunct neuroimaging, or simply the neurosurgical procedures and related hospitalization — undoubtedly raises the threshold for conducting these studies. Few are therefore willing to invest the time, effort and funds required to conduct such trials without robustly convincing, smaller-scale, preliminary studies. Even

studies that are designed ostensibly as feasibility and safety studies will often be evaluated on the basis of likely efficacy in order to justify the cost of proceeding to the next phase of trials, regardless of whether that justification is assessed by government, industry, academic or philanthropic interests. Unfortunately, this tendency — to use underpowered studies as a basis to move forward, or not, with larger, more definitive trials to assess clinical efficacy — may in fact add more noise than signal to the process of identifying and pursuing potentially useful therapies. This is because negative results are potentially falsely negative due to the underpowered nature of the preliminary studies, but even positive results may be falsely positive depending on the number of conditions tested and the unknown, underlying proportion of truly effective treatment conditions.

As an example, suppose a DBS feasibility study were designed to assess the effects of high vs. low frequency stimulation at a particular target for intractable, debilitating obsessive compulsive disorder, and the endpoints evaluated were reduction of obsessions and/or compulsions. The study was calculated to have a positive predictive power of 0.8 and results were to be accepted as statistically significant at $p < 0.05$. Suppose further that the unknown, true effect of DBS at this target for this condition was that only low frequency stimulation would be effective, and only for obsessions but not compulsions. Therefore, the proportion of true positive effects in this trial would be 1/4. Overall, then, what is the likelihood the results of this trial would reflect the underlying reality? A true positive would be revealed in 80% of cases. However, because only 1 in 4 conditions assessed were truly effective, a false positive would be detected in 14.3% of cases (resulting from the application of an alpha level of 0.05 to the three ineffective conditions). Therefore, combining the false negative rate (20%) and the false positive rate (14.3%) results in a trial that produces results that are misaligned with reality with a

probability approaching 1/3 (less than the simple sum of 0.20 and 0.143 because these events are not mutually exclusive, so some outcomes would overlap). This is, in essence, an extrapolation of the multiple-comparisons problem to clinical trials that cannot know in advance this so-called "base-rate" of true positive effects across the tested conditions.

These concerns regarding the overall validity of a result are especially relevant in the case of early clinical studies that, although explicitly directed towards establishing only feasibility or safety, nevertheless will often include an efficacy endpoint. For that purpose, they typically have relatively low power and might explore a wider range of parameters and outcomes to assess the broader potential of a particular therapeutic strategy. Despite their statistical and structural limitations, these studies are often evaluated for a "signal" of benefit, and decisions to proceed with larger trials may hinge upon these early efficacy results even if that is explicitly not the main purpose of those studies. Nevertheless, a clear understanding of the statistical limitations of these smaller, early-stage studies may suggest that the combined likelihood of false positives and false negatives is sufficiently high that putting too much emphasis on any result related to an underpowered, over-explored endpoint may be not much more reliable than flipping a coin.

From a statistical perspective, therefore, limiting the number of manipulations tested and the number of outcome measures assessed is desirable. However, in the field of neuromodulation, especially when applying electrical stimulation, this approach seems intuitively too restrictive: Given the large space of potential stimulation parameters and the complex, multi-dimensional nature of neuropsychiatric disease, selecting a small subset of protocols and outcomes may feel akin to blind spearfishing, whereas what we would like to do, ideally, is cast a wide net to discover a useful therapy.

Because ablation procedures have fewer degrees of freedom (i.e., no stimulation parameters to adjust), these interventions may seem a simpler and potentially more powerful neuromodulation technique in the context of clinical trials and may serve to guide and constrain the later development of stimulation techniques that build upon those lesion results. Nevertheless, lesions cannot capture the full set of effects achievable with various stimulation protocols, so one might be misled by false negatives if relying exclusively on predicate lesion studies to attempt neurostimulation. Furthermore, lesions, like electrical stimulation, vary in anatomical specificity and reproducibility; post-hoc analyses assessing outcomes as a function of exact lesion size or position, for example, will therefore be subject to similar potential multiple comparisons problems.

To limit the multiplicity of parameters related to the delivery and assessment of an investigational therapy, a neuromodulation study must be built upon a sound scientific premise. For example, if a particular neuroanatomical pathway is hypothesized to mediate a specific dysfunction such that its modulation might mitigate a related symptom, computational analysis such as finite element modeling of electrical fields and neuronal responses might yield a narrow set of stimulation parameters to be tested in order to produce the desired circuit effect (e.g., CENTURY-S: NCT02881151; ADvance II: NCT03622905). This, of course, presumes that at least type and direction of the desired neural response is understood. In other words, should the target or pathway be driven, inhibited, or recruited in some other manner to generate plasticity or release modulatory chemical factors? If this question cannot be answered with reasonable confidence, it may be a sign the planned study is premature.

One potentially interesting approach to sifting through the enormous space of potential neuromodulation protocols (e.g., electrical stimulation sites and patterns) is to include an

exploratory phase within the trial design, during which the goal is to achieve some effect on a well-defined biomarker hypothesized to mediate the intended therapeutic effect. For example, if fronto-medial theta power is proposed to influence depression [29], a flexible trial design could be implemented in which stimulation is "tuned" to produce the desired modulation of theta in that region within each patient, and that empirically-determined pattern of stimulation is then delivered continuously in the next phase of the trial to assess efficacy. This sort of adaptive trial design, when conducted according to a rigorously pre-specified plan, may accelerate progress towards a useful neuromodulation therapy [30].

Clarifying the Therapeutic Model

All studies implicitly or explicitly propose a particular causal structure to underlie potential interactions between variables, including the experimental manipulation, the outcome measures, and additional associated factors. While classical statistical methods were developed a tradition devoid of causality, work over the last 30 years has revealed the importance of designing studies and conducting analyses within a sound framework of putative causal interactions to minimize bias [31]. Constructing an explicit causal diagram for the proposed therapeutic model may provide useful clarity for the design of an appropriate study (Fig. 38.1). Such a diagram would lay bare the logic of the proposed investigation to facilitate a critical analysis of the plausibility of the scientific premise. Furthermore, this diagram would clearly identify mediating and confounding factors, such that the former might be used to derive secondary outcome measures while the latter are addressed with appropriate controls in design or analysis.

To highlight the importance of the proposed causal model in experimental design, suppose a nonrandomized, prospective pilot study is conducted to assess the effect of a novel neurostimulation strategy for a particular neuropsychiatric condition. All patients undergo stimulation and are followed so their outcomes can be compared to pre-operative baselines. Overall, a non-significant positive therapeutic trend is observed. To determine if there might be a subgroup of responders, patients expressing a putative biomarker that is postulated to enable or mediate the therapeutic effect — perhaps a particular neural rhythm or metabolic neuro-imaging alteration — undergo a subgroup analysis. In other words, this biomarker is hypothesized to be caused by the therapy and itself to be the cause the beneficial behavioral effect. However, suppose instead that, while stimulation does indeed tend to increase the expression of this biomarker, patients who happen to show improvements (e.g., due to placebo effects or other study activities) will also tend to show this biomarker independent of stimulation. In other words, observation of this biomarker may result from stimulation and/or from behavioral change, so it is in fact a colliding factor rather than a mediating one (as per Fig. 38.1). In this *post-hoc* analysis, even though there may be no causal pathway from stimulation to behavioral outcome, spurious associations between these can nevertheless be observed: if the threshold biomarker level used to select the subgroup is chosen such that either the stimulation or the behavioral response may be sufficient to cross it, a spurious negative correlation can be observed; if, however, the threshold is selected such that the combined (independent) influences of the stimulation and behavioral response are more likely to produce a supra-threshold level, then a spurious positive correlation between stimulation and behavioral outcome can be observed. In each case, the false association is termed a collider bias and results from the application of a threshold that screens out a group of non-stimulated *and* non-responding patients (in a manner analogous to Berkson's Paradox

[32]) in the context of a “true” causal link between the biomarker and behavioral outcome that is inverted with respect to the proposed therapeutic model.

This is but one of many possible examples that can demonstrate the variety of structural pitfalls related to designing and conducting clinical studies. Therefore, explicit elaboration of the proposed therapeutic model underlying a given experimental design and rigorous validation of the hypothesized causal steps within that model will decrease the likelihood of inferential errors.

Selecting Outcome Measures

The primary endpoint, or outcome measure, is ideally a thing of clear, undeniable value to quality and/or length of life. Unfortunately, there are relatively few real-world examples where endpoints are so simple, especially in the context of complex neuropsychiatric disease. For example, a new treatment that reduces a patient’s depression according to standard scales (e.g., MADRS or HAM-D) but fails to improve overall social and economic function could be regarded as a success or failure, depending on one’s viewpoint. Part of the confusion arises from the fact that those standard scales are, fundamentally, surrogate rather than true endpoints. A true primary endpoint would reflect what a patient desires from a therapy (leaving aside the thorny issues that arise when patients are unable to convey those desires or when they lack insight into their own needs). However, such desires will be heterogeneous and poorly quantifiable, so the use of surrogate measures is in fact the rule rather than the exception. In other words, our goal is to improve the lives of our patients, however they may imagine that improvement in the context of their illnesses (“I want to be able to do my woodworking again”), but a clinical study must necessarily homogenize individual variation through the selection or design of appropriate

surrogate measures. In many domains, particular surrogates have, for better or worse, become accepted standards for assessing therapeutic success (e.g., UPDRS in Parkinson's Disease).

Despite the fact that surrogate endpoints are far more common than typically appreciated, there exists debate surrounding their proper use [33]. This debate is related to the use of surrogates for the final outcome measure (a surrogate for a surrogate, in the framework presented here) and typically arises from the misconception that a viable surrogate outcome is simply any outcome that correlates well with the "true" outcome [34]. However, an ideal surrogate measure is one that fully predicts the effect of a treatment on the true outcome, and may in fact deterministically mediate the effect of treatment on that outcome. In practice, there are many available analytical methods to assess the validity of a proposed surrogate, each with particular strengths, weaknesses, and ideal application scenarios [35].

When secondary outcomes are fashioned in order to support the causal chain of a particular therapeutic model, the informative value of a trial can be greatly enhanced [36]. If, for example, both the primary and mediating secondary outcomes are met, the validity of a successful primary outcome is rendered more plausible. Conversely, when both primary and secondary outcomes are not achieved, one learns that either the therapeutic model is simply incorrect, or the model is correct but a failure to engage early mechanisms prevented the success of the primary outcome. Meanwhile, failed secondary outcomes with successful primary outcomes suggest the model is incorrect, or perhaps the primary outcome's success was spurious. Lastly, a failed primary outcome with successful mediating secondary outcomes suggests the therapeutic model may be incomplete or incorrect, or perhaps the failed primary outcome represented a false negative. For those with experience writing computer code, this process is

akin to debugging a function by reporting out the intermediate states of key variables as the code runs.

Designing Appropriate Control Conditions

Unlike most drug trials, neuromodulation trials cannot simply administer a placebo to a control group. Rather, the “placebo” in the case of many surgical trials is typically some sort of sham procedure [37]. Neurosurgical neuromodulation presents the possibility of additional forms of control: The control condition could take the form of sham surgery in lesion and infusion studies, placebo-delivery in infusion studies, or maintaining some group of implanted patients in a blinded, non-stimulated state in neuro-stimulation studies. The latter can take the form of delayed-start or withdrawal protocols, or cross-over designs.

Neurosurgical sham studies, in particular, are fraught with difficulties. Aside from the basic concern that some patients will undergo an invasive procedure that cannot be expected to provide any benefit beyond a likely transient placebo effect, the threshold regarding what constitutes “too little” or “too much” in a sham procedure is difficult to determine. For example, in a lesion study, if a sham lesion involved inserting a radio-frequency or laser probe directly into the target structure, one could argue that a micro-lesion effect may persist and mimic to some degree the actual lesion, and therefore this is not a true control, but a partial treatment. Conversely, if the probe is not inserted fully to target, those control patients will have less of the transient effects of probe insertion at the target (e.g., due to edema), and so will not represent a true control that differs only in the actual creation of a lesion; any improvement observed in the treatment group, especially if transient, could therefore reflect these “insertion effects.”

The AAV2-GAD study [9, 38] to assess the efficacy of gene therapy delivered to the subthalamic nucleus of patients with Parkinson's Disease is an example of a sham-controlled drug infusion study which engendered some disagreement related to the limited extent of sham surgery employed. The control procedure involved a partial thickness burr hole without the insertion of the drug delivery catheter into the brain. This study was criticized for the lack of any possibility of a micro lesion effect, which is commonly observed at this target in DBS surgery. Much of the criticism arose from the fact that infusion studies, unlike lesion studies, have the opportunity to perform a more rigorous control procedure, in which all patients undergo catheter placement and infusion, but controls receive only vehicle. A difference in effect can therefore more cleanly be ascribed to the putatively active compound. Practically speaking, however, it was almost certainly better from the perspective of fulfilling enrollment requirements to tell patients that, were they randomized to the control procedure, nothing would be inserted or infused into the brain; yet even this limited sham procedure might be regarded as too much by some patients [39].

More recently, the availability of focused ultrasound (FUS) may offer the potential for more acceptable sham procedures, as in the pivotal trial of FUS for essential tremor (ET) [8]. In that trial, no ultrasound energy was delivered to the control patients, who nevertheless underwent a head-shave, targeting scans, and in-scanner physical testing. However, to the extent that patients might have understood that FUS creates immediate lesions and that thalamotomy should be associated with improvements in tremor, this sham procedure may not have been truly blinded given the strong precedents and expectations. The only variable being tested, really, was whether the FUS device can indeed produce a sufficiently accurate thalamotomy lesion, and at least some

patients would likely have presumed there was enough evidence of this technical capability to run the trial in the first place.

From a regulatory perspective, this trial was sufficient to garner FDA approval for this method of treating medically intractable ET. From a clinical perspective, however, the primary question patients now ask is whether FUS or DBS is the “better” treatment, however they define that term. To many in the field, given that FUS can produce focal lesions, and given the knowledge that VIM thalamotomy is an effective lesion procedure for ET, the success of this trial in simply reducing tremor was not surprising or of primary interest. Rather, assessments of effect size and durability, as well of related complications (e.g., dysarthria, ataxia, persistent paresthesias, etc.), as compared to other neurosurgical treatments, particularly DBS, were the main concerns. One might therefore argue that once the technical capabilities of the FUS device have been established, given the known efficacy of thalamotomy for ET more generally, a sham-controlled trial of FUS thalamotomy was not likely to be very informative, and what was really needed was a comparison with other methods of surgically treating ET. In situations where the baseline efficacy of a particular lesion procedure remains uncertain (e.g., psychiatric disease, where randomized, sham-controlled lesion studies are rare and sometimes indeterminate [40]), this approach might have been more valuable. For ET, should a head-to-head trial of DBS vs. FUS have been performed, and could it have been performed in a rigorous fashion to examine both the baseline efficacy of the new procedure and to compare it to the existing DBS option? Should FUS have been compared to other lesion procedures, rather than to a sham procedure or to DBS? Furthermore, practically speaking, who would pay for such a head-to-head comparison? In many ways, the executed trial was the simplest and cleanest, even though the most pressing

clinical questions were not directly addressed. For now, meta-analyses of separate trials are the only available means of comparison [41].

The lesson here is that the “ideal” control may depend on one’s perspective. From an industry or commercial viewpoint, the simplest design to achieve regulatory approval is desirable. From a scientific standpoint, however, the comparison between treatment and control groups should inform clinically-relevant decisions. These goals do not necessarily align in every case.

Neurostimulation trials may, in many respects, be the simplest to control. Many recent studies have employed the strategy of implanting every subject, and then comparing “OFF” to “ON” conditions, whether that is between subjects (e.g., staggered start or withdrawal of active stimulation in one group) or within subjects (cross-over design). Which of these methods to employ may depends on the type of benefit the stimulation is anticipated to provide. For example, if neurostimulation is expected to provide a simple symptomatic benefit (e.g., tremor reduction), a cross-over design may be ideal, because each patient serves as her/his own control (in addition to being able to compare ON and OFF conditions across patients); if there is a benefit, all patients can eventually be placed in the ON state once the trial is complete. Alternatively, if stimulation is predicted to provide a disease-modifying benefit, particularly one that may accumulate with time, then it may be worthwhile to follow patients for longer durations and avoid turning the system off once it has been turned on. In this case, a delayed start paradigm might be more optimal, because the group that is initially ON can continue to be monitored for cumulative effect on disease progression over time (rather than have them ON then OFF then later back ON which could potentially start the accumulation process over again), as was done in the ADvance trial of fornix DBS for Alzheimer’s Disease [42, 43].

Any of these neurostimulation study designs assumes, for the sake of blinding, that the patients cannot distinguish the ON vs. OFF states. For this reason, it may be worthwhile including a formal experiment in these studies in which patients are challenged in a 2-alternative forced-choice paradigm to distinguish whether the stimulator is active or inactive at different settings. This could serve as an important guide to determine which settings are consistent with the goal of double-blinded ON vs. OFF assessments.

Selecting Anatomical Targets

Identifying the optimal anatomical target(s) is a core problem in neuromodulation studies, and the factors involved in this decision are at least as diverse as the conditions to be treated. Here, general principles that may be relevant across a broad range of studies are considered.

Ideally, there should be evidence of a causal link between activity in the target structure or pathway and disease manifestations. Nevertheless, co-variance of activity at the target and disease expression may be helpful when grounded in a well-supported circuit model; even if an area is downstream of those brain regions that directly cause symptom expression, there is the possibility of a retrograde influence of electrical stimulation, for example. Whether retrograde, anterograde, or on-target stimulation is more effective or specific is likely to vary by target and disease.

In principle, a combination of these approaches (e.g., multi-site stimulation to affect multiple nodes in a targeted circuit) could augment efficacy or specificity of a particular stimulation effect; one could imagine that specificity should be enhanced if a lower level of stimulation were applied to multiple nodes in a single network, such that local off-target effects would be reduced, but the effective perturbation applied to that system would nonetheless reflect

an approximate sum of that applied to the individual sites. Lesions, of course, may be synergistic as well. For example, combined cingulotomy and subcaudate tractotomy — the so-called “limbic leukotomy” — appeared to benefit OCD patients who had previously undergone only cingulotomy without significant improvement [44].

This notion that neuromodulation targets are properly networks rather than individual anatomical sites has gained increasing attention and acceptance, especially within the realm of epilepsy surgery [45-48]. While epilepsy may be somewhat an outlier in terms of the strength and coherence of the underlying neural activity, the premise that any behavior, be it normal or symptomatic, arises from the concerted activity of a distributed set of connected neural structures is not controversial [49]. What is not known is whether addressing multiple nodes for neuromodulation will consistently improve therapeutic outcomes, or whether there are cases where single targets are not only sufficient but perhaps optimal. In some cases, there may be a direct relationship between the complexity of disease manifestations and the corresponding size of the optimal target network. For example, most “functional” neurosurgeons have probably wondered if combined STN and GPi DBS in Parkinson’s Disease might provide synergistic benefits, given the incomplete overlap between the sets of therapeutic benefits observed at each target [5, 50-52]. Perhaps this strong degree of therapeutic overlap itself is reflective of the monosynaptic relationship between these target structures. A corollary of the idea that distributed neuromodulation may benefit multiple dimensions of a disease state may be that the span of benefit across dimensions is related to the network distance (number of synapses) and strength of connectivity between targeted sites.

The degrees of freedom that any particular neuroanatomical target affords — that is, the number of distinct dimensions of behavior that can be differentially modified — may be related

to its place in the phylogenetic hierarchy. In other words, more complex behaviors evolved more recently and are likely mediated by more recently appearing structures. Thus brainstem stimulation of ascending neuromodulatory systems may function as a fairly simple gain control on certain aspects of arousal, attention, reinforcement or decision thresholds [53-56], whereas cortical stimulation would be expected to produce more complex effects on particular domains of behavior. In fact, given the significant heterogeneity of responses seen in the cortex, especially areas such as the prefrontal cortex where single neurons heterogeneously encode complicated mixtures of task-specific features [57], addressing the full space of information processing available at those sites is likely to require much more fine-scale control of specific cortical columns or layers than is currently possible, though there is certainly interest in developing that capability (e.g., by companies such as Neuralink and Kernel).

White matter targets provide a means to address a potentially wide area of cortex (or other connected structures) within a small volume. Ventral internal capsule lesions or stimulation is one example of this approach, in which the goal is to modify the functioning of the orbitofrontal cortex broadly in patients with intractable, debilitating obsessive compulsive disorder [58]. Such a strategy, however, necessarily gives up hope of achieving precise behavioral effects with a high degree of specificity. Nonetheless, nonspecific effects over a broad domain of behavior may in fact be what are needed for therapeutic effect in some disease states. Conversely, the apparently limited cognitive sequelae of white matter lesions, particularly within the frontal lobes, can seem quite surprising. Cingulotomy, ventral capsulotomy and subcaudate tractotomy generally result in relatively subtle cognitive changes [59-64], and may even improve performance in some cognitive domains [65]; these likely reflect the distributed complexity and plasticity of the frontal lobes, as well as the potential ability of information to be routed around

the lesion (e.g., while a capsulotomy severs many subcortical orbitofrontal connections, cortico-cortical connections are left intact). However, whether this serves as some simple gain control on the overall function of that region, or instead is affecting some subtle aspect of its function in a general manner is unknown (e.g., in capsulotomy, does the orbitofrontal cortex simply have less access to reinforcement mechanisms via the cortical-basal ganglia-thalamic-cortical loop, but nevertheless retains the ability to process information from other cortical areas much as it had before?).

These considerations reveal that our ability to identify optimal neuroanatomical targets for neuromodulation in specific disorders is limited in large part by our rudimentary understanding of circuit-level disease mechanisms and a nascent understanding of normal systems-level function in the relevant brain areas. Even once a candidate target is identified, the manner in which its activity is to be modified poses yet another significant challenge. Is the goal to block activity to mimic lesion-type benefits? Is the goal to normalize rhythms that serve as "carrier" signals for effective neural processing? Is stimulation intended to "bridge" damaged or dysfunctional circuits, as in many brain-machine interface type projects? What types of stimulation will achieve the desired neural effects? These questions, though beyond the scope of this discussion, are clearly central to the development of new neuromodulation therapies.

Selecting Patients

Two factors are most relevant to defining the most appropriate patient population for neuromodulation within a selected disease entity: included patients should represent a relatively prevalent and typical form of the disease, and those patients should represent a stage or manifestation of the disease that is amenable to modification, either symptomatically or

pathophysiologically. In addition, selection criteria should ideally be fairly straightforward to deploy, such that if the neuromodulation therapy were found to be effective, continued success in broader application would not be limited by lack of necessary tools to identify appropriate patients (e.g. ultra-high-field MRI, expensive and sparsely available molecular testing, etc.) or by complex patient selection protocols that yield a mismatch between the studied population and the actually-utilized population.

In PD, for example, it is well-known that patients who are beyond a certain stage of disease progression are ill-suited for DBS [66]; the motor benefits are unlikely to improve overall quality of life significantly because dementia and other non-motor factors have become the major problems, and DBS may even exacerbate dementia directly. So the extent to which the disease in such patients is "modifiable," especially from the motor perspective, is limited. The "ADvance" trial of fornix DBS for Alzheimer's Disease was designed with the premise that earlier intervention would lead to better outcomes, so patients with mild, "probable" Alzheimer's dementia were recruited [43]. This study, however, included patients with early-onset dementia who were atypical of the overall disease state. Post-hoc analyses suggested these patients may have negatively biased the DBS effect (though of course such post-hoc analyses are to be viewed with caution given the large number of potential comparisons that can be applied after-the-fact). The ensuing phase of that trial, therefore, raised the minimum age of entry to be better aligned with the typical Alzheimer's population (ADvance II: NCT03622905).

Three major studies of DBS for depression have been conducted, one targeting the subcallosal cingulate [12] another targeting the ventral striatum and ventral portion of the anterior limb of the internal capsule [13], and a third targeting a similar region of the anterior limb of the internal capsule [18]. The former two trials failed to achieve overall efficacy

endpoints, although there were individual cases where positive treatment effects were clearly evident (for example, in which sudden re-emergence of symptoms was found to be related to an unappreciated depletion of the pulse generator battery). These failures are possibly attributable to incomplete specification of the anatomical target, which therefore varied by patient [67], as well as to suboptimal patient selection. Meanwhile, patients were enrolled based upon diagnosis of depression according to criteria in the Diagnostic and Statistical Manual (DSM), which applies a checklist of fungible diagnostic criteria, summing the number of checked items to surpass a somewhat arbitrary threshold. Therefore, patients with different constellations of symptoms may be assigned the same overall diagnostic label. Importantly, patients with depression are known to manifest symptoms in particular clusters that may have distinct neurobiological mechanisms [68]. Therefore, subtype heterogeneity may have confounded the results of these depression studies. This type of phenomenon has been observed in OCD, where particular subtypes (e.g., "hoarders") appear less likely to benefit from capsulotomy [69]. For these reasons, if subgroups are known to exist but there is no *a priori* hypothesis regarding which ones are likely to benefit, adequately powering the study for pre-specified sub-group analyses may be a worthwhile effort, or limiting the study to a particular subgroup might be considered.

The notion that DSM disease categories may be insufficient to capture distinct neuropathological entities has led to the movement to characterize behavior according to traits that are potentially more fundamental as seen from a neural systems perspective. Specifically, the U.S. National Institutes of Health "Research Domain Criteria" (RDoC) framework views behavior through the lens of six domains (positive valence, negative valence, cognitive systems, social processes, arousal/regulatory systems, and sensorimotor systems). Each domain contains constructs that focus on particular functions (e.g., sensitivity to reward and ability to use reward

for learning, in the case of positive valence systems). Behavioral performance on a battery of tasks that assess these functions is hypothesized to be a more sensitive and specific classifier of neuropsychological function than traditional criteria [70]. While the particular tasks employed to assess these functions are not nearly as standardized or validated as common clinical tools (i.e., DSM categories, standard scales and routine neuropsychological tests), the promise of this "first principles" approach to behavior is that neuropsychiatric disease might be defined according to underlying mechanisms, and this would certainly be more desirable for neuromodulation trials.

Determining Frequency and Duration of Follow-Up

Neuromodulation studies, like all clinical trials, must balance a desire for frequent and long-term follow-up with practical considerations such as burden on patients, cost, and timely reporting of results. In many cases, patients will travel long distances to participate in these studies, and so the financial and logistical hurdles resulting from frequent, ongoing follow-up can be significant. Nevertheless, there is a clear need to maximize data collection to benefit the overall reliability of results.

Most neuromodulation studies include some set of behavioral assessments at pre-specified intervals. In light of the high variability of behavioral performance even within subjects across sessions, days and months, more frequent measurement will serve to reduce noise. Because the power of a study is inversely related to the variability of the measurements, to the extent that more accurate assessments of individual subjects are obtainable (e.g., taken as the average over more frequent estimates), it may be possible to use this increased signal to noise as a means to counterbalance the number of subjects required for the study. There may in fact be an overall cost savings because additional assessments of patients who have already undergone an

expensive neuromodulation-related procedure are almost certainly less expensive than performing more procedures on additional patients, even if travel, housing and other related expenses are required.

Because many neuromodulation therapies show gradual symptomatic responses, a sufficient duration of follow-up can be critical to the proper evaluation of a new approach. One prominent example of an inadequate duration to primary endpoint leading to apparent therapeutic failure is the SANTE trial of thalamic DBS for epilepsy [23]. Here, seizure reduction at the pre-specified 3-month endpoint did not reach the threshold for success, so FDA approval was not immediately obtained. Nevertheless, clear improvements in seizure burden were observed at six months and beyond [71]. Eventually, based upon these extended data, regulatory approval was granted. Like epilepsy, dystonia and OCD are also known to exhibit a gradually improving response to neuromodulation with time [72, 73], so future trials are likely to benefit by keeping these examples in mind when scheduling final endpoints.

Planning for Post-Protocol Patient Care

Neurostimulation studies are distinct in that subjects are implanted with complicated electronic devices that are often intended to be permanent. When a study ends, these devices nonetheless remain. These devices may require ongoing maintenance (replacement of implanted batteries, replacement or repair of wireless charging equipment, etc.). If a study has succeeded and regulatory approval is granted, long-term care may not be a problem. More likely, however, when a study fails, the status of ongoing care for the device may become uncertain with respect to the responsible parties and the cost.

While one might presume naively that a failed trial should not provide any reason to continue support for an ineffective device, patients often feel differently. In some cases, there truly are individual patients who are receiving benefit from the implanted device; in such cases, withdrawal of device support after a patient has undergone the risks and troubles of having the device implanted and participating in the study could be viewed as unethical. Yet even if there is no clear evidence that a patient is receiving benefit, many patients will perceive benefit — whether through the filter of a placebo effect or because they are truly experiencing something not captured in the study protocol — or they may at least continue to hold onto the hope of benefit. In these cases, many might conclude the most ethical course of action is to continue device support. Even warning patients during the consent process that if the study fails device support will not continue may not be a sufficient means to avoid this ethical responsibility because, at the very least, there may in fact be a real individual benefit.

Committing to and budgeting for long-term device support is therefore the best approach, whenever possible. Device manufacturers have been generally helpful in providing continued support after a failed trial. Insurance companies tend to be more heterogeneous in their willingness to pay for related procedures, but often this is possible if a clinical team is sufficiently persistent and able to work through the logistical hurdles. Depending on the extent to which this becomes a more widespread problem as the number of neuromodulation trials increases, perhaps competitors within academia and industry would be willing to work together to expand the field as a whole by creating a fund, fed by a tax on individual trials, to support at least the medical costs incurred by ongoing device maintenance when other means to cover those costs (insurance or philanthropy) are unavailable.

Conclusions

The advancement of neuromodulation will be driven by successful clinical studies. The success of these studies, however, does not hinge upon solely the achievement of the primary therapeutic goals, but rather derives equally from their ability to deliver confident answers to well-posed questions. We should realistically expect most trials to fail in achieving their predicted therapeutic outcomes; if that were not the case, it would reflect a lack of ambition to overcome the most difficult and pressing problems faced by individuals with neuropsychiatric disease. Designing trials that are just as informative in failure as they are in success is necessary to ensure progress in the field as well as to justify the risks undertaken by the brave and pioneering patients who enroll in these studies.

Trials embarked upon too early, without sufficient rationale, not only expose patients to potentially unnecessary risks, but threaten overall progress. Likewise, a well-reasoned neuromodulation strategy examined via a suboptimal study design may fail to reveal the truth of that strategy. Although interest in neuromodulation has expanded greatly as technological ability and neuroscientific knowledge have improved, there is nevertheless a finite set of resources, both in terms of economic and human capital, to apply towards sufficiently intensive and large-scale studies. Therefore, optimizing trial design by heeding the positive and negative lessons of those studies preceding is essential to realize the brightest future of neuromodulation.

References

1. Teskey GC, Monfils MH, VandenBerg PM, Kleim JA. 2002. Motor map expansion following repeated cortical and limbic seizures is related to synaptic potentiation. Cereb Cortex 12: 98-105

2. van Rooyen F, Young NA, Larson SE, Teskey GC. 2006. Hippocampal kindling leads to motor map expansion. Epilepsia 47: 1383-91

3. Kupsch A, Benecke R, Muller J, Trottenberg T, Schneider GH, et al. 2006. Pallidal deep-brain stimulation in primary generalized or segmental dystonia. N Engl J Med 355: 1978-90

4. Deuschl G, Schade-Brittinger C, Krack P, Volkmann J, Schäfer H, et al. 2006. A randomized trial of deep-brain stimulation for Parkinson's disease. The New England journal of medicine 355: 896-908

5. Odekerken VJJ, van Laar T, Staal MJ, Mosch A, Hoffmann CFE, et al. 2013. Subthalamic nucleus versus globus pallidus bilateral deep brain stimulation for advanced Parkinson's disease (NSTAPS study): a randomised controlled trial. Lancet neurology 12: 37-44

6. Weaver FM, Follett K, Stern M, Hur K, Harris C, et al. 2009. Bilateral deep brain stimulation vs best medical therapy for patients with advanced Parkinson disease: a randomized controlled trial. JAMA : the journal of the American Medical Association 301: 63-73

7. Follett KA, Weaver FM, Stern M, Hur K, Harris CL, et al. 2010. Pallidal versus subthalamic deep-brain stimulation for Parkinson's disease. The New England journal of medicine 362: 2077-91

8. Elias WJ, Lipsman N, Ondo WG, Ghanouni P, Kim YG, et al. 2016. A Randomized Trial of Focused Ultrasound Thalamotomy for Essential Tremor. N Engl J Med 375: 730-9

9. LeWitt PA, Rezai AR, Leehey MA, Ojemann SG, Flaherty AW, et al. 2011. AAV2-GAD gene therapy for advanced Parkinson's disease: a double-blind, sham-surgery controlled, randomised trial. Lancet Neurol 10: 309-19

10. Marks WJ, Jr., Bartus RT, Siffert J, Davis CS, Lozano A, et al. 2010. Gene delivery of AAV2-neurturin for Parkinson's disease: a double-blind, randomised, controlled trial. Lancet Neurol 9: 1164-72

11. Lang AE, Gill S, Patel NK, Lozano A, Nutt JG, et al. 2006. Randomized controlled trial of intraputamenal glial cell line-derived neurotrophic factor infusion in Parkinson disease. Ann Neurol 59: 459-66

12. Holtzheimer PE, Husain MM, Lisanby SH, Taylor SF, Whitworth LA, et al. 2017. Subcallosal cingulate deep brain stimulation for treatment-resistant depression: a multisite, randomised, sham-controlled trial. Lancet Psychiatry 4: 839-49

13. Dougherty DD, Rezai AR, Carpenter LL, Howland RH, Bhati MT, 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: 240-8

14. Mallet L, Polosan M, Jaafari N, Baup N, Welter ML, et al. 2008. Subthalamic nucleus stimulation in severe obsessive-compulsive disorder. N Engl J Med 359: 2121-34

15. Lopes AC, Greenberg BD, Canteras MM, Batistuzzo MC, Hoexter MQ, et al. 2014b. Gamma ventral capsulotomy for obsessive-compulsive disorder: a randomized clinical trial. JAMA psychiatry 71: 1066-76

16. Lozano AM, Fosdick L, Chakravarty MM, Leoutsakos JM, Munro C, et al. 2016b. A Phase II Study of Fornix Deep Brain Stimulation in Mild Alzheimer's Disease. J Alzheimers Dis 54: 777-87

17. Denys D, Mantione M, Figee M, van den Munckhof P, Koerselman F, et al. 2010. Deep brain stimulation of the nucleus accumbens for treatment-refractory obsessive-compulsive disorder. Arch Gen Psychiatry 67: 1061-8

18. Bergfeld IO, Mantione M, Hoogendoorn ML, Ruhe HG, Notten P, et al. 2016. Deep Brain Stimulation of the Ventral Anterior Limb of the Internal Capsule for Treatment-Resistant Depression: A Randomized Clinical Trial. JAMA Psychiatry 73: 456-64

19. Wiebe S, Blume WT, Girvin JP, Eliasziw M, Group EaEoSfTLES. 2001. A randomized, controlled trial of surgery for temporal-lobe epilepsy. The New England journal of medicine 345: 311-18

20. Heck CN, King-Stephens D, Massey AD, Nair DR, Jobst BC, et al. 2014. Two-year seizure reduction in adults with medically intractable partial onset epilepsy treated with responsive neurostimulation: final results of the RNS System Pivotal trial. Epilepsia 55: 432-41

21. Barbaro NM, Quigg M, Ward MM, Chang EF, Broshek DK, et al. 2018. Radiosurgery versus open surgery for mesial temporal lobe epilepsy: The randomized, controlled ROSE trial. Epilepsia 59: 1198-207

22. Barbaro NM, Quigg M, Broshek DK, Ward MM, Lamborn KR, et al. 2009. A multicenter, prospective pilot study of gamma knife radiosurgery for mesial temporal lobe epilepsy: seizure response, adverse events, and verbal memory. Annals of neurology 65: 167-75

23. Fisher R, Salanova V, Witt T, Worth R, Henry T, et al. 2010. Electrical stimulation of the anterior nucleus of thalamus for treatment of refractory epilepsy. Epilepsia 51: 899-908

24. North RB, Kidd DH, Farrokhi F, Piantadosi SA. 2005. Spinal cord stimulation versus repeated lumbosacral spine surgery for chronic pain: a randomized, controlled trial. Neurosurgery 56: 98-106; discussion 06-7

25. Kumar K, Taylor RS, Jacques L, Eldabe S, Meglio M, et al. 2007. Spinal cord stimulation versus conventional medical management for neuropathic pain: a multicentre randomised controlled trial in patients with failed back surgery syndrome. Pain 132: 179-88

26. Kapural L, Yu C, Doust MW, Gliner BE, Vallejo R, et al. 2015. Novel 10-kHz High-frequency Therapy (HF10 Therapy) Is Superior to Traditional Low-frequency Spinal Cord Stimulation for the Treatment of Chronic Back and Leg Pain: The SENZA-RCT Randomized Controlled Trial. Anesthesiology 123: 851-60

27. Penn RD, Savoy SM, Corcos D, Latash M, Gottlieb G, et al. 1989. Intrathecal baclofen for severe spinal spasticity. N Engl J Med 320: 1517-21

28. Brown JA, Lutsep HL, Weinand M, Cramer SC. 2006. Motor cortex stimulation for the enhancement of recovery from stroke: a prospective, multicenter safety study. Neurosurgery 58: 464-73

29. Broadway JM, Holtzheimer PE, Hilimire MR, Parks NA, Devylder JE, et al. 2012. Frontal theta cordance predicts 6-month antidepressant response to subcallosal cingulate deep brain stimulation for treatment-resistant depression: a pilot study. Neuropsychopharmacology 37: 1764-72

30. Chow SC. 2014. Adaptive clinical trial design. Annu Rev Med 65: 405-15

31. Mansournia MA, Higgins JP, Sterne JA, Hernan MA. 2017. Biases in Randomized Trials: A Conversation Between Trialists and Epidemiologists. Epidemiology 28: 54-59

32. Porta M, Vineis P, Bolumar F. 2015. The current deconstruction of paradoxes: one sign of the ongoing methodological "revolution". Eur J Epidemiol 30: 1079-87

33. Weintraub WS, Luscher TF, Pocock S. 2015. The perils of surrogate endpoints. Eur Heart J 36: 2212-8

34. Alonso A, Van der Elst W, Molenberghs G, Buyse M, Burzykowski T. 2015. On the relationship between the causal-inference and meta-analytic paradigms for the validation of surrogate endpoints. Biometrics 71: 15-24

35. Molenberghs G, Burzykowski T, Alonso A, Assam P, Tilahun A, Buyse M. 2010. A unified framework for the evaluation of surrogate endpoints in mental-health clinical trials. Stat Methods Med Res 19: 205-36

36. Landau S, Emsley R, Dunn G. 2018. Beyond total treatment effects in randomised controlled trials: Baseline measurement of intermediate outcomes needed to reduce confounding in mediation investigations. Clin Trials 15: 247-56

37. Probst P, Grummich K, Harnoss JC, Huttner FJ, Jensen K, et al. 2016. Placebo-Controlled Trials in Surgery: A Systematic Review and Meta-Analysis. Medicine (Baltimore) 95: e3516

38. Niethammer M, Tang CC, LeWitt PA, Rezai AR, Leehey MA, et al. 2017. Long-term follow-up of a randomized AAV2-GAD gene therapy trial for Parkinson's disease. JCI Insight 2: e90133

39. Cohen PD, Isaacs T, Willocks P, Herman L, Stamford J, et al. 2012. Sham neurosurgical procedures: the patients' perspective. Lancet Neurol 11: 1022

40. Lopes AC, Greenberg BD, Canteras MM, Batistuzzo MC, Hoexter MQ, et al. 2014a. Gamma ventral capsulotomy for obsessive-compulsive disorder: a randomized clinical trial. JAMA Psychiatry 71: 1066-76

41. Harary M, Segar DJ, Hayes MT, Cosgrove GR. 2019. Unilateral Thalamic Deep Brain Stimulation Versus Focused Ultrasound Thalamotomy for Essential Tremor. World Neurosurg

42. Leoutsakos J-MS, Yan H, Anderson WS, Asaad WF, Baltuch G, et al. 2018. Deep Brain Stimulation Targeting the Fornix for Mild Alzheimer Dementia (the ADvance Trial): A Two Year Follow-up Including Results of Delayed Activation. Journal of Alzheimer's disease : JAD 64: 597-606

43. Lozano AM, Fosdick L, Chakravarty MM, Leoutsakos J-M, Munro C, et al. 2016a. A Phase II Study of Fornix Deep Brain Stimulation in Mild Alzheimer's Disease. Journal of Alzheimer's disease : JAD 54: 777-87

44. Bourne SK, Sheth SA, Neal J, Strong C, Mian MK, et al. 2013. Beneficial effect of subsequent lesion procedures after nonresponse to initial cingulotomy for severe, treatment-refractory obsessive-compulsive disorder. Neurosurgery 72: 196-202; discussion 02

45. Bartolomei F, Lagarde S, Wendling F, McGonigal A, Jirsa V, et al. 2017. Defining epileptogenic networks: Contribution of SEEG and signal analysis. Epilepsia 58: 1131-47

46. Bernhardt BC, Bonilha L, Gross DW. 2015. Network analysis for a network disorder: The emerging role of graph theory in the study of epilepsy. Epilepsy Behav 50: 162-70

47. Laxpati NG, Kasoff WS, Gross RE. 2014. Deep brain stimulation for the treatment of epilepsy: circuits, targets, and trials. Neurotherapeutics 11: 508-26

48. Spencer SS. 2002. Neural networks in human epilepsy: evidence of and implications for treatment. Epilepsia 43: 219-27

49. Stam CJ. 2014. Modern network science of neurological disorders. Nat Rev Neurosci 15: 683-95

50. Baizabal-Carvallo JF, Roze E, Aya-Kombo M, Romito L, Navarro S, et al. 2013. Combined pallidal and subthalamic nucleus deep brain stimulation in secondary dystonia-parkinsonism. Parkinsonism Relat Disord 19: 566-8

51. Boel JA, Odekerken VJJ, Schmand BA, Geurtsen GJ, Cath DC, et al. 2016. Cognitive and psychiatric outcome 3 years after globus pallidus pars interna or subthalamic nucleus deep brain stimulation for Parkinson's disease. Parkinsonism & related disorders 33: 90-95

52. Weaver FM, Follett KA, Stern M, Luo P, Harris CL, et al. 2012. Randomized trial of deep brain stimulation for Parkinson disease: thirty-six-month outcomes. Neurology 79: 55-65

53. Avery MC, Krichmar JL. 2017. Neuromodulatory Systems and Their Interactions: A Review of Models, Theories, and Experiments. Front Neural Circuits 11: 108

54. Carter ME, Yizhar O, Chikahisa S, Nguyen H, Adamantidis A, et al. 2010. Tuning arousal with optogenetic modulation of locus coeruleus neurons. Nat Neurosci 13: 1526-33

55. Lee SH, Dan Y. 2012. Neuromodulation of brain states. Neuron 76: 209-22

56. Lin SC, Brown RE, Hussain Shuler MG, Petersen CC, Kepecs A. 2015. Optogenetic Dissection of the Basal Forebrain Neuromodulatory Control of Cortical Activation, Plasticity, and Cognition. J Neurosci 35: 13896-903

57. Rigotti M, Barak O, Warden MR, Wang X-J, Daw ND, et al. 2013. The importance of mixed selectivity in complex cognitive tasks. Nature

58. Rasmussen SA, Norén G, Greenberg BD, Marsland R, McLaughlin NC, et al. 2017. Gamma Ventral Capsulotomy in Intractable Obsessive-Compulsive Disorder. Biological Psychiatry

59. Kartsounis LD, Poynton A, Bridges PK, Bartlett JR. 1991. Neuropsychological correlates of stereotactic subcaudate tractotomy. A prospective study. Brain 114 (Pt 6): 2657-73

60. Nyman H, Andreewitch S, Lundback E, Mindus P. 2001. Executive and cognitive functions in patients with extreme obsessive-compulsive disorder treated by capsulotomy. Appl Neuropsychol 8: 91-8

61. Ochsner KN, Kosslyn SM, Cosgrove GR, Cassem EH, Price BH, et al. 2001. Deficits in visual cognition and attention following bilateral anterior cingulotomy. Neuropsychologia 39: 219-30

62. Ridout N, O'Carroll RE, Dritschel B, Christmas D, Eljamel M, Matthews K. 2007. Emotion recognition from dynamic emotional displays following anterior cingulotomy and anterior capsulotomy for chronic depression. Neuropsychologia 45: 1735-43

63. Subramanian L, Bracht T, Jenkins P, Choppin S, Linden DE, et al. 2017. Clinical improvements following bilateral anterior capsulotomy in treatment-resistant depression. Psychol Med 47: 1097-106

64. Williams ZM, Bush G, Rauch SL, Cosgrove GR, Eskandar EN. 2004. Human anterior cingulate neurons and the integration of monetary reward with motor responses. Nature Neuroscience 7: 1370-75

65. Batistuzzo MC, Hoexter MQ, Taub A, Gentil AF, Cesar RC, et al. 2015. Visuospatial Memory Improvement after Gamma Ventral Capsulotomy in Treatment Refractory Obsessive-Compulsive Disorder Patients. Neuropsychopharmacology 40: 1837-45

66. Daniels C, Krack P, Volkmann J, Pinsker MO, Krause M, et al. 2010. Risk factors for executive dysfunction after subthalamic nucleus stimulation in Parkinson's disease. Mov Disord 25: 1583-9

67. Riva-Posse P, Choi KS, Holtzheimer PE, Crowell AL, Garlow SJ, et al. 2018. A connectomic approach for subcallosal cingulate deep brain stimulation surgery: prospective targeting in treatment-resistant depression. Mol Psychiatry 23: 843-49

68. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, et al. 2017. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat Med 23: 28-38

69. Gentil AF, Lopes AC, Dougherty DD, Rück C, Mataix-Cols D, et al. 2014. Hoarding symptoms and prediction of poor response to limbic system surgery for treatment-refractory obsessive-compulsive disorder. Journal of Neurosurgery 121: 123-30

70. Cuthbert BN, Insel TR. 2013. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Med 11: 126

71. Salanova V, Witt T, Worth R, Henry TR, Gross RE, et al. 2015. Long-term efficacy and safety of thalamic stimulation for drug-resistant partial epilepsy. Neurology 84: 1017-25

72. Kupsch A, Tagliati M, Vidailhet M, Aziz T, Krack P, et al. 2011. Early postoperative management of DBS in dystonia: programming, response to stimulation, adverse events, medication changes, evaluations, and troubleshooting. Mov Disord 26 Suppl 1: S37-53

73. Rasmussen SA, Noren G, Greenberg BD, Marsland R, McLaughlin NC, et al. 2018. Gamma Ventral Capsulotomy in Intractable Obsessive-Compulsive Disorder. Biol Psychiatry 84: 355-64

74. Pearl J. 2010. An introduction to causal inference. Int J Biostat 6: Article 7

Table 38.1. A non-exhaustive list of major neuromodulation clinical trials enrolling at least 15 subjects, conducted in a prospective, randomized fashion. The last column, success, relates to successful achievement of the primary outcome.

Name of Study	**n**	**Design**	**Primary Outcomes**	**Success ?**
Movement Disorders				
Pallidal Deep-Brain Stimulation in Primary Generalized or Segmental Dystonia [3]	40	All patients implanted with DBS in GPi, randomized to active or sham-stimulation (no stimulation delivered) for 3 months, followed by 3-6 months of open-label treatment	Burke-Fahn-Marsden Dystonia Rating Scale change from baseline - 3 months	yes
STN-Stimulation Versus Best Medical Treatment in Advanced PD [4]	156	Unblinded 1:1 randomization to stimulation or best medical management	Parkinson's Disease Questionnaire (PDQ-39, quality of life), Unified Parkinson's Disease Rating Scale (UPDRS-III), baseline to six-months	yes
Subthalamic nucleus versus globus pallidus bilateral deep brain	12	Patients 1:1 randomized to STN vs. GPi DBS; patients and assessors	Baseline to 12 months: AMC linear disability scale (ALDS), reliable	no

stimulation for advanced Parkinson's disease (NSTAPS study) [5]	8	blinded to target.	change index (RCI), mini-international neuropsychiatric interview (MINI), UPDRS	
A Comparison of Best Medical Therapy and Deep Brain Stimulation of Subthalamic Nucleus and Globus Pallidus for the Treatment of Parkinson's Disease [6]	255	Patients 1:1 randomized to best medical therapy or DBS, DBS patients additionally randomized/split to GPi or STN; motor evaluation performed by blinded neurologists	Baseline to six months: Time spent in the 'on' state w/o dyskinesias (by motor diaries)	yes
CSP #468 Phase II - A Comparison of Best Medical Therapy and Deep Brain Stimulation of Subthalamic Nucleus and Globus Pallidus for the Treatment of Parkinson's Disease [7]	299	Patients 1:1 randomized to STN or GPi target; DBS neurologists blinded to target	Baseline to 24 months: change in UPDRS-III	no
ExAblate Transcranial MR Guided	76	Patients 3:1 randomized to unilateral	Clinical Rating Scale for Tremor and	yes

Focused Ultrasound for the Treatment of Essential Tremors [8]		HIFU thalamotomy or sham procedure	the Quality of Life in Essential Tremor Questionnaire at 3 months post-op	
Study of AAV-GAD Gene Transfer Into the Subthalamic Nucleus for Parkinson's Disease [9]	45	Patients 1:1 randomized to sham surgery or AAV2-GAD infusions	UPDRS part III at 6 months post-op	yes
Double-Blind, Multicenter, Sham Surgery Controlled Study of CERE-120 in Subjects With Idiopathic Parkinson's Disease [10]	51	Patients 1:1 randomized to AAV2-NRTN infusions or sham surgery	UPDRS part III at 15 months post op	no
Randomized controlled trial of intraputamenal glial cell line–derived neurotrophic factor infusion in Parkinson disease [11]	34	Patients 1:1 randomized to glial cell line-derived neurotrophic factor or saline infusion	UPDRS part III at 6 months post-op	no
Psychiatric Disorders				
Subcallosal cingulate deep brain	90	All patients implanted with DBS in	>= 40% reduction in depression	no

stimulation for treatment-resistant depression: a multisite, randomized, sham-controlled trial [12]		bilateral subcallosal cingulate white matter, randomized to 6 months of active (n=60) or sham (n=30) DBS, followed by 6 months of open-label DBS	severity from baseline	
A Randomized Sham-Controlled Trial of Deep Brain Stimulation of the Ventral Capsule/Ventral Striatum for Chronic Treatment-Resistant Depression [13]	30	All implanted in ventral capsule/ventral striatum, 1:1 randomized to active vs sham DBS treatment in a blinded fashion for 16 weeks, then open-label phase	>= 50% improvement on Montgomery Åsberg Depression Rating Scale from baseline at 16 weeks	no
STOC Study: Subthalamic Nucleus Stimulation in Severe Obsessive–Compulsive Disorder [14]	17	Patients 1:1 randomized to on-off or off-on stimulation, 3 months each with a 1-month washout period in the middle, double-blind	Yale-Brown Obsessive Compulsive Scale at the end of two 3-month periods	yes

Radiosurgical Treatment for Obsessive-compulsive Disorder [15]	16	Patients 1:1 randomized to gamma ventral capsulotomy or sham surgery. Patients blinded for 1 year post-op	Yale-Brown Obsessive Compulsive Scale at 1 year post-op	no
ADvance trial: Deep brain stimulation of the fornix for early, probable Alzheimer's Disease [16]	42	All implanted in fornix, 1:1 randomized to active and sham stimulation for first 12 months, all patients active the following year	ADAS-cog (Alzheimer's disease assessment scale - cognitive component), Clinical Dementia Rating sum of boxes, cerebral glucose metabolism measured with PET	no
Deep brain stimulation of the nucleus accumbens in treatment-refractory patients with obsessive-compulsive disorder [17]	16	All patients implanted, stimulated for 8 months, 1-month double-blind (2 weeks on stimulation, 2 weeks off stimulation)	Yale-Brown Obsessive Compulsive Scale at each 2-week double-blind interval	yes
Deep brain stimulation of the ventral anterior limb of the internal capsule for depression [18]	25	All patients implanted, 16 randomized to OFF then ON vs. ON then OFF (cross-over design, each phase lasting 2-3 weeks)	> 50% reduction in Hamilton-D 17-item scale	yes
Epilepsy				

A Randomized, Controlled Trial of Surgery for Temporal-Lobe Epilepsy [19]	80	Patients 1:1 randomized to surgery or best medical therapy	Freedom from seizures that impair awareness of self and surroundings	yes
RNS System Pivotal trial: Responsive Neurostimulation for Epilepsy [20]	191	All patients implanted to 1 or 2 foci; patients 1:1 randomized to active or sham stimulation 1 month post-op; evaluated 3 months later	Seizure frequency	yes
Radiosurgery versus open surgery for mesial temporal lobe epilepsy: The randomized, controlled ROSE trial [21]	58	Patients 1:1 randomized to stereotactic radiosurgery or anterior temporal lobectomy; evaluating neurologists were blinded to procedure	Self-reporting of seizure frequency between 25 and 36 months post-op	yes

A multicenter, prospective pilot study of gamma knife radiosurgery for mesial temporal lobe epilepsy: Seizure response, adverse events, and verbal memory [22]	30	Patients 1:1 randomized to 20 or 24Gy targeting the amygdala, hippocampus, and parahippocampal gyrus	Self-reporting of seizure frequency at 36 months post-op	yes
SANTE: Stimulation of the Anterior Nucleus of the Thalamus for Epilepsy [23]	110	All implanted in anterior nucleus of thalamus, 1:1 assigned to active or no stimulation for first 3 months, months 4-13 were unblinded	reduction in monthly seizure rate after 3 months	yes
Spine				
Spinal Cord Stimulation versus Repeated Lumbosacral Spine Surgery for Chronic Pain: A Randomized, Controlled Trial [24]	60	Patients 1:1 randomized to lumbosacral spine reoperation or spinal cord stim	>= 50% pain relief, patient satisfaction, reoperation at six months	yes
Spinal cord stimulation versus conventional medical management for neuropathic pain: A	100	Patients 1:1 randomized to spinal cord stim or conventional medical	>= 50% leg pain relief at six months	yes

multicentre randomised controlled trial in patients with failed back surgery syndrome [25]		management (patients not blinded)		
Comparison of 10-kHz High-Frequency and Traditional Low-Frequency Spinal Cord Stimulation for the Treatment of Chronic Back and Leg Pain [26]	198	Patients 1:1 randomized to conventional spinal cord stim (~50 Hz) or 10 kHz spinal cord stim	>= 50% back pain reduction, no stimulation related neurological deficit at three months	yes
Intrathecal Baclofen for Severe Spinal Spasticity [27]	20	Patients received 2x3-day alternating trials of saline or baclofen	Muscle tone with Ashworth score at end of baclofen 3-day period	yes
Stroke				
Everest Trial: Epidural Electrical Stimulation for Stroke Rehabilitation [28]	164	Patients 2:1 randomized to implanted epidural motor cortex stimulation or control (no sham surgery); all patients underwent same schedule of rehabilitation; evaluating clinicians	upper-extremity Fugl-Meyer (UEFM) and arm motor ability test (AMAT) at 4 weeks post-rehabilitation	no

		blinded to treatment		

Figure Legend

Fig. 38.1. A causal diagram depicting the types of variables that may be present in any particular neuromodulation trial design. A randomized trial, by assigning the experimental manipulation in a stochastic fashion that is not subject to any causal inputs other than the "flip of a coin," in principle removes the possibility of confounding factors. Whether this is true in practice, of course, depends on the size of the studied population and the distributions of characteristics across groups; a happenstance clustering of particular features such as age, gender, disease severity or subtype, etc., can undermine the randomization. Mediating factors are causally related to the treatment effect, whereas their proxies are related to that effect only insofar as they directly correspond to those mediating factors themselves; secondary outcomes reporting proxies of mediating factors are therefore valid only to the extent of that correspondence. Note that downstream effects can be used as proxies for the behavioral outcomes, subject to the same type of constraint. Colliding factors are present when a factor has multiple potential causes, and here are depicted as caused by both the manipulation and the assessed outcome (behavioral state). In non-randomized, observational studies these are often mistaken for confounding factors, and post-hoc stratification of outcomes by "controlling for" these factors can result in spurious correlations between hypothesized treatment and effect [74]. Note that this sort of "collision" also occurs at the behavioral outcome, resulting from causal inputs via the manipulation and non-confounding factors. Stratification of outcomes to assess relationships among potentially relevant experimental variables can result in spurious correlations between the manipulation and these non-confounding factors, similar to the situation with colliding factors, because behavioral state is, here, technically a colliding factor as well.