Reporting Guidelines and Issues to Consider for Using Intracranial Brain Stimulation in Studies of Human Declarative Memory

Nanthia Suthana 1,2,3*, Zahra M. Aghajan 1, Emily A. Mankin 2 and Andy Lin 4

1 Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, Jane and Terry Semel Institute for Neuroscience and Human Behavior, UCLA, Los Angeles, CA, United States, 2 Department of Neurosurgery, David Geffen School of Medicine, UCLA, Los Angeles, CA, United States, 3 UCLA, Los Angeles, CA, United States, 4 IDRE Statistical Consulting Group, UCLA, Los Angeles, CA, United States

Participants with stimulating and recording electrodes implanted within the brain for clinical evaluation and treatment provide a rare opportunity to unravel the neuronal correlates of human memory, as well as offer potential for modulation of behavior. Recent intracranial stimulation studies of memory have been inconsistent in methodologies employed and reported conclusions, which renders generalizations and construction of a framework impossible. In an effort to unify future study efforts and enable larger meta-analyses we propose in this mini-review a set of guidelines to consider when pursuing intracranial stimulation studies of human declarative memory and summarize details reported by previous relevant studies. We present technical and safety issues to consider when undertaking such studies and a checklist for researchers and clinicians to use for guidance when reporting results, including targeting, placement, and localization of electrodes, behavioral task design, stimulation and electrophysiological recording methods, details of participants, and statistical analyses. We hope that, as research in invasive stimulation of human declarative memory further progresses, these reporting guidelines will aid in setting standards for multicenter studies, in comparison of findings across studies, and in study replications.

Keywords: declarative memory, intracranial stimulation, deep brain stimulation, humans, medial temporal lobe

INTRODUCTION

The use of surgically implanted electrodes within the human brain has become increasingly common in treating and/or evaluating abnormal brain activity in patients with epilepsy (Schulze-Bonhage, 2017), Parkinson’s disease (Benabid et al., 1987), and dystonia (Vidalhiet et al., 2005) and is also being explored in depression (Ressler and Mayberg, 2007), obsessive compulsive disorder (Nuttin et al., 1999), and Alzheimer’s disease (Lozano et al., 2016). With a continued rise in medical treatments using implanted neural devices, research opportunities to record from, and stimulate, the brain during human cognition will also likely increase. For the study of human declarative memory, it is common to work with participants with temporal lobe epilepsy (TLE), as the temporal lobe plays an important role in forming and retrieving declarative memories (i.e., facts and events). In TLE, seizures are thought to originate from a site within the temporal lobe, and electrodes are therefore either placed on the surface—such as in lateral temporal
cortex—or implanted deeply within medial temporal lobe (MTL) regions, making research studies on declarative memory modulation using direct brain stimulation possible.

TLE patients with implanted electrodes are relatively rare in a research setting, and thus single-site intracranial stimulation studies of declarative memory frequently suffer from low sample sizes (e.g., <10). Even if multi-site studies can claim large samples (e.g., >20) through data collection at multiple surgical centers, numerous factors such as electrode placement, characteristics (e.g., diameter) and implantation approaches often vary within these samples, resulting in small homogeneous subsamples used for statistical analyses. An important goal of these studies should therefore be to report findings in a way that will allow for accurate evaluation, replication, and future meta-analyses. Since the use of intracranial stimulation for memory modulation is still a relatively small field, the goal of this mini-review is to provide a set of reporting guidelines that will facilitate future comparisons across studies and proper replication. We hope these guidelines will generate productive discussions between fellow researchers and provide a framework to guide design and/or evaluation of similar studies. Similar efforts have proven to be productive in other research fields such as functional magnetic resonance imaging (Poldrack et al., 2008). A summary of relevant studies is presented in Table 1, guidelines are discussed below and a reference checklist is provided in Appendix A.

**DESCRIBE THE PARTICIPANTS FULLY**

The ability to probe the human brain and apply electrical stimulation to investigate cognitive functions is a unique and an invaluable opportunity. It must be noted, however, that these experiments are largely done in patients with epilepsy and consequently, there is a concern regarding the generalizability of results to the normal population. Thus, when interpreting the results, there are certain confounding factors that cannot be eliminated but must be explicitly mentioned and taken into consideration.

Furthermore, because this data acquisition is often difficult and time consuming, published findings usually have few participants within a given statistical group. To facilitate comparisons across studies and ease their integration into future larger meta-analyses, detailed information of study participants should be provided. In addition to basic demographic information, studies should report relevant clinical information such as medications, neuropsychological scores, MRI abnormalities, comorbid neuropsychiatric conditions, and determined seizure onset zone (SOZ). It has been shown that interictal epileptic activity can interfere with cognitive performance (Kleen et al., 2013; Ung et al., 2017), so when possible, a quantification of the frequency of such events should be included.

One important consideration is that the targeting of implanted electrodes is determined based on clinical criteria. This implies that some implanted electrodes, potentially including the stimulating electrode, can fall within the SOZ. Stimulating the SOZ can lead to afterdischarges or other nonspecific effects (Cherlow et al., 1977; Kesner, 1982; Halgren and Wilson, 1985) that can interfere with results. It is, thus, necessary to be cognizant of any interactions between stimulation effects on memory and seizure activity. If stimulating the SOZ cannot be avoided, one possibility is to compute statistical tests using data from all stimulating electrodes, and then confirm whether results are consistent when considering only data from when the stimulating electrode was in clinically determined “healthy tissue.” A discrepancy could indicate that stimulation in the SOZ differentially affects memory processes. Any relationship between memory effects and seizure activity or proximity of electrode to SOZs should be reported. At minimum, the determined SOZ, occurrence of seizures, and/or seizure related activity should be reported for each participant.

Yet another factor in conducting memory research in epilepsy patients is that participants can have impaired cognitive functions. This can potentially influence the efficacy of stimulation on memory modulation. Thus, it is important to include participant level information about cognitive abilities. Ideally, results reported will be pertinent to the type of memory tested with stimulation. Examples of relevant neuropsychological tests reported in previous studies include: Wechsler Memory Scale (verbal memory) (Wechsler, 2005); California Verbal Learning Test (verbal memory) (Delis et al., 2000); Rey–Osterrieth Complex Figure Test (visual memory) (Meyers and Meyers, 1995).

Clinical studies using DBS in Parkinson’s disease have highlighted the effects medication can have on neurophysiological activity and responsiveness to stimulation (Brown and Williams, 2005). While clinical circumstances in epilepsy patients do not allow for researchers to control the presence of medication, studies should report details of medications that participants take, including the type of medication and, ideally, the time of most-recent administration relative to study completion.

**REPORT IN DETAIL ON ELECTRODE CHARACTERISTICS AND ELECTRODE LOCALIZATION METHODS**

Recent clinical DBS studies have emphasized the importance of precise electrode location in treatment outcome. The precise position of electrodes within white or gray matter could be critical for efficacy of treatment in both Parkinson’s disease and depression (Pouratian et al., 2011; Riva-Posse et al., 2014). Implanted electrodes in TLE patients are targeted to specific regions based on clinically hypothesized SOZs in each individual. Therefore, studies often contain substantial variability among electrode locations. While subdural (i.e., strip and grid) electrodes lie on the surface of the brain and affect gray matter areas, depth electrodes can stimulate white and/or gray matter, which can have different effects on behavioral outcomes (Titiz et al., 2017). Furthermore, areas like the amygdala consist of distinct nuclei where stimulation location may be critical (Inman et al., 2018). Providing high-resolution MRI
| Study                         | Type of memory                     | Memory impairment or improvement | Area(s) of stimulation | Unilateral/ Bilateral | Electrode type | Electrode properties | Electrode implantation approach |
|------------------------------|------------------------------------|----------------------------------|------------------------|-----------------------|-----------------|----------------------|---------------------------------|
|                             |                                    |                                  |                        |                       |                 |                      |                                  |
| Halgren et al., 1985         | Complex scenes (recognition)       | Impairment                       | MTL                    | Bilateral             | Depth           | NR                   | NR                              |
| Halgren and Wilson, 1985     | Word pairs (recall)                | Impairment                       | MTL                    | Unilateral            | Depth           | NR                   | NR                              |
| Fernández et al., 1996       | Words                              | Impairment                       | Hippocampus            | Unilateral            | Depth           | NR                   | ~1 mm                           |
| Coleshill et al., 2004       | Words, objects, faces (recognition)| Impairment                       | Hippocampus            | Unilateral            | Depth           | NR                   | NR                              |
| Hamani et al., 2008          | Word pairs (recognition)           | Improvement                       | Fornix                 | Bilateral             | Depth           | 1.27 mm              | NR                              |
| Lacruz et al., 2010          | Words, objects, faces (recognition)| Improvement                       | Rhinal cortex          | Bilateral             | Depth           | 1.3 mm              | NR                              |
| Fell et al., 2013            | Words (recall and recognition)     | Improvement                       | Rhinal cortex and hippocampus | Bilateral  | Depth | 1.5 mm | NR |
| Coleshill et al., 2012       | Spatial navigation                 | Improvement                       | Entorhinal             | Unilateral            | Depth           | 1.5 mm              | NR                              |
| Koubelis et al., 2013        | Mini-mental status exam            | Improvement                       | Fornix                 | Unilateral            | Depth           | 1.1 mm              | NR                              |
| Miller et al., 2015          | Complex figures (recall)           | Improvement                       | Fornix                 | Unilateral            | Depth           | 1.1 mm | NR | |
| Jacobs et al., 2016          | Words (recall), spatial navigation | Improvement                       | MTL                    | Unilateral            | Depth | area = 0.099 cm$^2$ | NR | |
| Titiz et al., 2017           | Persons (recognition)              | Improvement                       | Entorhinal             | Unilateral            | Depth           | 100 µm              | NR                              |
| Merkow et al., 2017          | Words (recall)                     | Improvement                       | MTL                    | Unilateral            | Depth           | 1.2 mm              | NR                              |
| Ezzyat et al., 2017          | Words (recall)/ Improvement        | Improvement                       | Variable               | Unilateral            | Depth           | 2.4 mm              | NA                              |
| Kuczewicz et al., 2018a      | Words (recall)                     | Improvement                       | Lateral temporal       | Unilateral            | Depth           | NR                  | NR                              |
| Kuczewicz et al., 2018b      | Words (recall)                     | Improvement                       | Lateral temporal       | Unilateral            | Depth           | 1–10 mm             | NR                              |
| Ezzyat et al., 2018          | Words (recall)                     | Improvement                       | Lateral temporal       | Unilateral            | Depth           | 0.8–2.3 mm          | NR                              |
| Kim et al., 2018             | Spatial navigation, temporal order | Improvement                       | Lateral temporal       | Unilateral            | Depth           | 0.8–2.3 mm          | NR                              |
| Hansen et al., 2017          | Word-color pairs (recognition)     | No effect                         | Entorhinal             | Unilateral            | Depth           | 1.3 mm              | NR                              |
| Irnman et al., 2018          | Objects (recognition)              | Improvement                       | Amygdala               | Unilateral            | Depth           | 0.86 mm             | NR                              |

(Note: (Continued)
| Study | Bi/Monopolar | Amplitude | PW  | Frequency  | Charge density | Impedance | Stimulation timing | Neurophysiological changes | AED | SOZ  |
|-------|-----------|-----------|-----|-----------|----------------|-----------|-------------------|--------------------------|-----|------|
| Halgren et al., 1985 |
| Halgren and Wilson, 1985 |
| Fernández et al., 1996 |
| Colshill et al., 2004 |
| Hamani et al., 2008 |
| Lacruz et al., 2010 |
| Fell et al., 2013 |
| Suthana et al., 2012 |
| Koubeissi et al., 2013 |
| Miller et al., 2015 |
| Jacobs et al., 2016 |
| Titz et al., 2017 |
| Merkov et al., 2017 |
| Ezzyat et al., 2018a |
| Kucewicz et al., 2018a |
| Kucewicz et al., 2018b |
| Ezzyat et al., 2018b |

(Continued)
Studies should report in detail how electrode contact locations are determined; namely, the type of registration procedure used and how electrode contacts and brain regions are visualized.

Studies should include, at the minimum, an electrode localization figure for an example participant in the main analyses (e.g., Figure 1) and an electrode localization figure or table showing the placement of electrode contacts for each participant, perhaps in the Supplemental Materials. Each participant’s electrode localizations should be included with a unique subject ID that is consistent throughout the manuscript, allowing localization information to be cross-referenced to other information (e.g., behavioral or electrophysiological results) regarding individual subjects. If stimulation or electrophysiological analyses are done using bipolar montages, localizations should be reported for each individual electrode contact. Since accurate localization of electrodes depends on the quality of both the post-implantation (e.g., CT) scans and pre-implantation (e.g., MRI) scans, studies should report acquisition parameters, and consequent voxel resolution of all scans to enable accurate evaluation and future replication. Furthermore, in addition to detailed description of registration procedures, the known minimal error associated with the procedure should be reported. Subdural electrodes provide an additional challenge for localizations, as brain shift frequently accompanies the implantation procedure. Thus, studies should either correct for that shift or acquire high resolution post-implantation MRI scans (Groppe et al., 2017).

Several toolboxes are available for demarcating subcortical anatomy in individual subjects to aid in visualization of small subregions in which electrode contacts reside (e.g., Yushkevich et al., 2010). The toolbox used and any relevant parameter settings should be reported. Given the differences between the electrical properties of gray and white matter, and because electrodes may not lie neatly inside a single tissue type, it is helpful to report an estimation of relative amount of gray and white matter in the vicinity of each stimulating electrode (Mercier et al., 2017).

Electrode locations visualized on an average template brain could also be included to demonstrate patterns of electrode distribution across participants using known toolboxes (e.g., Xia et al., 2013). However, non-linear registration and inter-subject anatomical variability can cause aggregate figures to lose valuable within-subject detail that may help explain variability in findings across the sample and across studies from different surgical sites. With areas such as the MTL, where subregions can be millimeters in thickness, these subtle differences could be detrimental to replication efforts if individual localization information is not presented as well. Therefore, group brain visualizations of electrodes should not be the only method of electrode localization but rather supplemental to individual subject electrode localizations. Furthermore, registration procedures from individual to group or template brain images must be reported in detail.
Some studies include participants with different types of electrodes with varying diameter and/or spacing between contacts. For example, subdural electrodes can have up to 10 mm of space between adjacent stimulating bipolar contacts. Depth electrodes, which penetrate the MTL can vary, usually with 3–10 mm between contacts. Given the large spacing between electrode contacts used for bipolar stimulation, a pair of stimulating contacts rarely falls within the same area. Most studies, nonetheless, refer to one electrode location in a single brain area (e.g., only one contact or the calculated midpoint between two bipolar contacts), but it is not always clear how the reported area was chosen. We recommend reporting the location of each contact when bipolar stimulation is used. Indeed, providing this information could aid tremendously in the comparison of research studies and replication of methodologies across sites.

**FIGURE 1** | An example image for studies to follow when reporting single participant electrode localizations for bipolar macro-stimulation or monopolar micro-stimulation. (A) Coronal view of a co-registered image of a pre-implantation high-resolution CT and post-implantation high-resolution MRI with overlaid electrode contact segmentations. (B) The outcome of the electrode localization procedure where individual electrode contact locations within the left hippocampus are overlaid onto a high-resolution MRI for final visualization and reporting. Red circles: bipolar macro-stimulation contacts; red crosshair: monopolar microstimulation contact.

**DESCRIPTION BEHAVIORAL TASK DESIGN**

Studies should report in detail the behavioral task design and why it was selected. Ideally, enough details should be reported such that independent researcher would be able to replicate the task. Studies should also report if there were any differences in the behavioral design across subjects such as difficulty level (e.g., number of stimuli presented) and if so, account for this variable in their statistical analyses. Details of any unique circumstances for a given participant should be reported. How behavioral performance was calculated should be explained in detail and raw performance metrics for each condition should be reported—ideally for each individual subject—not only changes in performance between conditions (e.g., stimulation vs. non-stimulation).

**REPORT ALL PARAMETERS OF STIMULATION**

The electric field generated by intracranial stimulation electrodes is thought to govern the neural response to stimulation (Butson et al., 2007). As such, there are many variables that can modify the generated electric field and, thus, lead to different electrophysiological and behavioral outcomes.

**Amplitude and Impedance**

The amplitude of stimulation is controlled either in voltage or current domain, and the domain chosen should be reported. Although voltage is a critical factor, in particular, for determining the volume of tissue affected (Butson et al., 2007), current-controlled stimulation protocols have been implemented recently (Preda et al., 2016). Whether there are differences in clinical outcomes or therapeutic advantages of each type of stimulation is debated (Bronstein et al., 2015; York and Moro, 2017). Impedance of the stimulating electrode should also be measured and reported. This is important for participant safety reasons, but additionally, the reporting of impedance in combination with either voltage or current allows for translating to the other domain, thus enabling meta-analyses to choose a consistent measure.

**Frequency**

Varying frequencies of stimulation have been shown to affect the outcome of DBS studies in domains such as Parkinson’s disease (Moreau et al., 2008), dystonia (Kupsch et al., 2003), and depression (Mayberg et al., 2005). No intracranial study to date has systematically investigated the relationship between different stimulation frequencies and declarative memory performance.
Many studies have used continuous stimulation protocols for which reporting frequency and amplitude is sufficient. Animal studies have shown that theta-burst stimulation is especially effective for inducing long-term potentiation (Larson et al., 1986) and could, thus, be beneficial for memory. Recently, theta-burst stimulation has been used in humans (Miller et al., 2015; Titiz et al., 2017; Inman et al., 2018; Kim et al., 2018), and in such studies frequency of bursts must be reported along with frequency within each burst.

**Charge Density**

For stimulation to have an effect, it is necessary to deliver efficacious amount of charge without compromising the electrochemical balance of the tissue, which can lead to potential safety issues (Rose and Robblee, 1990). Thus, stimulation charge density is an important factor to consider, which can be calculated from the combination of the following parameters: duration of the stimulation pulse (T); surface area of the stimulation electrode contact (or contacts if bipolar stimulation, A); and the amplitude (current) of the pulse (I), according to the following equation: $\rho = \frac{IT}{A}$.

In order to facilitate future replication and meta-analyses, all numbers related to the waveform of the stimulation pulse and charge density should be reported.

In addition, reports should include any unique patient-level effects; such as whether they were aware of stimulation or whether there were any consequent side effects stimulation that was unexpected.

**DETAIL TIMING OF STIMULATION RELATIVE TO MEMORY TASK**

The temporal specificity of stimulation, both in terms of its duration and its timing with respect to other behavioral task parameters, can introduce variability in study outcomes. For instance, stimulation applied in the hippocampus at encoding but not retrieval impairs memory (Lacruz et al., 2010). Additionally, applying intermittent stimulation in the nucleus basalis in adult monkeys enhances working memory, but continuous stimulation leads to memory impairment (Liu et al., 2017). Studies should, therefore, report the timing of stimulation relative to stimulus presentation and whether stimulation was applied during encoding, distraction, retrieval, or any other time periods.

**CHOOSE THE APPROPRIATE STATISTICAL MODEL**

Statistical model assumptions commonly violated in intracranial stimulation studies of declarative memory are the assumption of independence of observations and assumptions regarding the distribution of the outcome or the errors. Although studies generally use repeated measures designs, not all statistical analyses address the non-independence of observations inherent in these designs. Many reported statistical models assume independence of observations but have nonetheless included observations from the same participant. Violation of the independence assumption can result in biased parameter estimates, underestimated variability of model parameters, and overly optimistic $p$-values (Hox et al., 2017). Researchers should consider using methods that support estimation assuming non-independence within clusters of observations, such as mixed models. Mixed models can include random effects that simultaneously attempt to account for the non-independence within clusters and also quantify heterogeneity across clusters. Many standard regression models (e.g., linear, logistic) have been extended to mixed models to allow for random effects. Parametric statistical models (e.g., $t$-tests, ANOVA, regressions) make assumptions about the distributions of the dependent variables or the errors, the individual deviations from the population means. While classical ANOVA and linear regression assume normally distributed errors, logistic regression assumes that the outcome is binomially-distributed, appropriate for a success/failure variable. Logistic regression can be used to model the probability of successfully recalling a single item, and to estimate effects of predictors that vary at the item level (e.g., stimulation on/off condition, time since the last stimulation). Additionally, logistic regression accounts for differing numbers of trials across participants, which linear regression of aggregated mean probabilities cannot. Finally, logistic regression can be extended to mixed models including random effects to address non-independence of success/failure outcomes within clusters of observations.

**TRANSPARENTLY REPORT DETAILS OF THE STATISTICAL ANALYSIS**

Given the rarity of intracranial stimulation studies of declarative memory, transparency in reporting of all statistical analyses is crucially needed. Unfortunately, however, relevant details are often obscured with only the statistical model used and $p$-values reported. Below we outline three details that should be clearly specified to improve reader evaluation of the statistical analyses including (1) sample sizes, (2) effect sizes, and (3) the number of hypotheses tested and $p$-values estimated over the entirety of the statistical analysis. Omission of these details can result in improper evaluation and interpretation of findings and problems with study replication.

**Sample Sizes**

Sample sizes inform the reader about the replicability of the analysis and should be reported for each statistical analysis or model. Sample sizes can vary greatly within the same study but across analyses when subjects are used selectively in analyses or when some analyses use aggregated variables (e.g., means, sums). Intracranial stimulation studies of declarative memory commonly use a repeated-measures design where participants experience several encoding sessions, brain regions, and memory tests. While some studies have modeled the probability of recall of a single item (Ezzyat et al., 2018; Kim et al., 2018), others have modeled the mean probability
of recall of items aggregated at the level of session (Jacobs et al., 2016), brain location (Jacobs et al., 2016; Merkow et al., 2017), or even participant (Kucewicz et al., 2018a), at times within the same study. The burden on the reader to remember the sample sizes at each level of aggregation can be cumbersome, so every analysis presented should include sample size information.

**Effect Sizes**

Effect sizes provide an estimate of the magnitude of the effect, which significance tests and p-values do not. Significance tests may suggest whether observed changes in memory were likely to arise by chance, but the associated statistics do not directly estimate how large that change is and whether it is substantively meaningful. Readers should be given the opportunity to interpret effect sizes for themselves, so effect sizes should be reported and interpreted. Precision in estimation of effect sizes can be expressed with confidence intervals, giving the reader a range of effect sizes compatible with the data. Both unstandardized effect sizes (e.g., mean differences, regression coefficients), and standardized effect sizes (e.g., Cohen’s d, R²), are worth reporting. The shift of emphasis away from null hypothesis testing and toward effect estimation is being advocated in many fields to address problems of low replicability (Cumming, 2014).

**Number of Hypothesis Tests and P-Values**

Another potential source of low replicability is the practice of reporting only significant findings. A low p-value from a single, planned statistical model will generally be more persuasive than a single low p-value from 10 modifications of a statistical model. However, researchers generally present the final model as if it were the only one tested. As more hypotheses are tested, the probability of making an erroneous inference increases, and without accounting for the additional testing, the reader’s interpretation of reported p-values will generally be too optimistic. To combat misinterpretation of p-values, the American Statistical Association recommends reporting the number of hypotheses and all statistical analyses conducted throughout the study (American Statistical Association, 2016). Additionally, researchers should make appropriate adjustments to significance thresholds to account for multiple comparisons.

**ELECTROPHYSIOLOGY**

Oftentimes, in addition to the implanted stimulation electrode—which typically cannot record neural signals during stimulation or at all—additional electrode contacts nearby or in other brain areas capable of recording electrophysiological activity are present. This situation presents additional opportunities to analyze neural responses during the behavioral task in response to sensory input, as well as to evaluate the effects of stimulation on those responses. In this case, the locations of both the stimulating and recording electrodes should be reported following the guidelines detailed above.

It is outside the scope of this review to give thorough guidelines for reporting on electrophysiology, however the presence of stimulation artifact in recorded data must be dealt with prior to drawing conclusions. It is critical to report the method used for rejecting stimulation artifacts, and include figures of signals before and after data cleaning procedures (e.g., Basir-Kazeruni et al., 2017; O’Shea and Shenoy, 2018). Because many stimulation artifact rejection algorithms rely on the assumption that the artifact does not saturate the signal (Basir-Kazeruni et al., 2017; O’Shea and Shenoy, 2018), amplitude of the artifact and recording ranges should be reported.

In the case where electrophysiological recordings are conducted in subjects with epilepsy, the ability to record also allows for detailed analysis of epileptic discharges during the memory task. These epileptic spikes have been shown to interfere with memory and other cognitive performance (Kleen et al., 2013; Ung et al., 2017) and should be taken into account when analyzing the data. In particular, manuscripts should report how epileptic discharges were detected and how much data was affected and/or excluded based on these detections. Statistical analyses of main effects should take into account the prevalence and timing of interictal spikes and summary statistics of these variables should be included.

**CLOSED LOOP STIMULATION**

Most studies to date have been conducted using open-loop stimulation, which considers only external factors in determining when and whether to stimulate. However, increasingly, studies are able to record neural signals during behavioral tasks and then use internal factors of the neural state to determine the precise timing of stimulation, commonly referred to as closed-loop stimulation. Internal brain states can be used to determine the timing of stimulation [e.g., at a particular phase of an endogenous neural oscillation (Fell et al., 2013)] or to make a decision about whether or not to stimulate at all.

The challenge of closed-loop decision-making has been recently tackled using machine learning models: a set of stimulation-free trials with neural data and labels indicating subsequent memory performance was collected and used to train a model to recognize features that predict memory states for a particular participant, which in turn could drive stimulation decisions (Ezzyat et al., 2017, 2018).

It is important when reporting closed-loop stimulation studies to not only report all of the parameters required for open-loop stimulation, but also include detailed descriptions of the prediction model used to trigger stimulation, how the training set was collected and size of the training set, model performance [e.g., area under the receiver operant characteristic curve (AUC)], criterion for when stimulation was given/withheld, and elapsed time between collecting training data and testing closed-loop stimulation. Ideally, to demonstrate that the model produces relevant decisions at the time of an experiment, closed-loop experiments should include both test, and control trials that are collected within a single experimental session. In both types of trials, the model’s stimulation decision would be recorded, however stimulation would only be applied during the test trials. This would allow for gauging the model performance without the confound of the stimulation.
CONCLUSIONS

Intracranial stimulation studies of declarative memory have reported both enhancement and impairment of memory, yet the specific factors that give rise to these differences in memory modulation are unclear. What is evident, however, are numerous differences across study sites in methodologies, including but not limited to details of participants, behavioral task design, electrode characteristics, electrode placements, stimulation parameters, timing of stimulation, and statistical methods. While true replication may be difficult in this field due to the rarity of participants and difficulties of completing large sample studies that are consistent in various within-sample characteristics, the first step toward replication should be transparent reporting of methods and results. We therefore, propose a set of guidelines to reporting and issues to consider when completing future intracranial stimulation studies of declarative memory.

AUTHOR CONTRIBUTIONS

NS, ZM, EM, and AL wrote manuscript. NS, ZM, and EM contributed to the figure and table.

FUNDING

Supported by grants from the National Institute of Neurological Disorders and Stroke (NS103802 and NS058280) and the A. P. Giannini Foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2018.00905/full#supplementary-material
