Editorial

Recommendations for the Design and Analysis of *In Vivo* Electrophysiology Studies

Introduction

The Journal of Neuroscience has initiated a series of editorials on issues related to experimental design across several research areas. The first piece (http://www.jneurosci.org/content/38/14/3375) focused on human neuroimaging studies, and the current editorial addresses experimental issues that should be considered when designing and analyzing *in vivo* physiology studies. Taking these issues into account before data collection begins can avoid some common concerns that arise during peer review.

1. Defining sample size

There is a long tradition in neurophysiology of using the number of neurons recorded as the sample size ("*n*") in statistical calculations. In many cases, the sample of recorded neurons comes from a small number of animals, yet many statistical analyses make the explicit assumption that a sample constitutes independent observations. When multiple neurons are recorded from a single animal, however, either sequentially with a single electrode or simultaneously with multiple electrodes, each neuron's activity may not, in fact, be independent of the others. Thus, it is important for researchers to account for variability across subjects in data analyses.

In different types of experimental design, statistical comparisons may use the number of recorded neurons, the number of slices or culture dishes (for *in vitro* studies), or the number of individual animals used for recordings. Ideally, "*n*" could be the number of animals from which a standard set of recordings have been made (e.g., multiunit recordings or local field potentials [LFPs] obtained from a fixed position) in different treatment groups. This may be appropriate when individual animals undergo different behavioral paradigms or systemic drug treatments before recording neuronal activity. However, this may be impractical or inappropriate for studies in which large populations of neurons are recorded from a small number of precious or rare animals, such as non-human primates. In particular, when individual neurons have different functional properties, it is not appropriate to average across all cells recorded in a single animal. Finally, when the number of neurons recorded from each animal varies substantially, data from some animals may be weighted more heavily than others.

JNeurosci does not have an inflexible rule governing how authors should determine "*n*" but recommends some best practices:

- It is essential that authors describe clearly how many animals contributed to a sample and how many neurons were recorded from each individual animal.
- A rationale for the number of animals and number of neurons recorded should be provided.
- Authors should determine whether differences across individuals can account for variance in outcome variables derived from neural recordings.
- Subject identity should be included as a covariate in data analyses whenever possible.
- Power analyses should be used when possible. Alternatively, evaluation of the strength of the evidence (e.g., Bayes factor for a particular hypothesis) can be performed as data are collected without inflating the risk of false-positives (Dienes, 2016).

2. Considerations for single-unit and multiunit recordings

Many laboratories now routinely use large electrode arrays, or optical techniques, to record from many neurons simultaneously. Optical techniques allow recording from up to thousands of neurons simultaneously and can provide information about their anatomical organization. However, these techniques are frequently limited to recordings near the brain surface and tend to use reporters with slow kinetics that reflect action potentials only indirectly. Microelectrode arrays can also record from many neurons; however, because the individual electrode contact sites generally cannot be moved independently, the isolation of single neurons is not always as good as it is with traditional single-electrode techniques. Thus, each method has strengths and weaknesses that should be considered when designing experiments. We will focus on electrical recordings here.

The use of array recordings has also introduced changes in the types of signals that are recorded, specifically single-unit versus multiunit activity. Traditionally, in experiments using individual microelectrodes, single-unit action potentials were isolated meticulously and recorded over time. While large electrode arrays increase yield greatly, they generally do not allow for careful unit isolation during data collection. Thus, many modern datasets consist of mixtures of well-isolated single units along with multiunit activity. Moreover, multiunit activity can mean very different things: in some cases, multiunit activity represents mainly spikes from a single neuron that is not sufficiently well isolated. In other cases, multiunit activity involves specifying a threshold and accepting all signal deflections that exceed such a threshold. In the latter case, a multiunit signal may consist of many undifferentiated spike trains, along with noise.

This issue is increasingly addressed with sophisticated spike sorting methods. However, automated spike sorting returns results that are often difficult to validate. Indeed, spike amplitude and shape may not be stationary over time, such that the isolation and composition of neural signals may change during data collection. Some spike-sorting approaches are designed to track such changes, but these methods are not foolproof.

While nonstationarity is always a concern, unit-isolation quality affects some scientific questions more than others, and this should be considered. Detailed questions about neural coding may require well-isolated single neurons. However, emerging

analysis approaches that characterize population activity may be less sensitive than mixed single-unit and multiunit neural data. Justification should, however, be provided if this is assumed. It would clearly be valuable for the field to converge on a set of widely used metrics to quantify isolation of neural signals. Given these challenges, *JNeurosci* recommends that authors follow reporting practices that maximize transparency:

- It is essential that authors describe their full signal processing workflow, including the specific criteria for classifying neural signals as single units or multiunits.
- The processing steps and statistical criteria used in automated spike-sorting algorithms should be documented. These may include measures taken to assess the stationarity of signals and the stability of unit isolation.
- When many neurons are recorded simultaneously, metrics should be provided that convey the quality of isolation across the range of recorded cells.
- If relevant to the question, data should be reported separately for single units and multiunits (e.g., with different symbols in a scatter plot), and statistical analyses should consider whether findings are consistent across these types of signals.
- Whenever possible, particularly in studies with a small number of neurons recorded, all data from individual neurons should be provided, which is possible as Extended Data.

3. Recording of field potentials

Continuous recordings within neural tissue are often used to measure aggregate electrical activity within a brain structure, termed the LFP. LFPs are robust against small electrode movements and allow comparison of brain activity from the same location over weeks and months. Alternatively, surface electrodes can be used in electrocorticography to measure an intracranial EEG signal, which has better signal-to-noise ratio and spatial resolution than extracranial EEGs.

These population signals are often quantified as responses evoked by an experimental event or are analyzed in the spectral domain by decomposition into frequency bands. However, common to all signals of this type is that the precise neural generators are seldom known because they arise from many elements (synapses, action potentials, circuits). The nature of the signal recorded at a given location may also depend on the nature of the electrode used; for example, electrode impedance and geometry affect their spatial selectivity. Furthermore, these signals are often prone to electrical artifacts, such as contamination by muscle activity, head movement, or epileptiform discharges.

To facilitate the interpretation of LFP and electrocorticography signals, and to promote reproducibility, we suggest the following guidelines:

- The location of recording and reference electrodes (including laminar depth and brain atlas coordinates, if applicable) should be reported; where signals from multiple electrodes are recorded in the same preparation, the electrode spacing should be provided.
- Where signals are grouped across recording sites for statistical comparisons, an explanation should be given of why signals from distinct sites can be regarded as statistically independent.
- The nature of electrodes, such as their diameter and material properties, including impedance, should be reported.
- Methods for artifact rejection should be detailed, including objective and reproducible criteria for excluding individual

trials or electrodes; the use of additional methods for monitoring potential artifacts (e.g., electrocardiograms, videotaping, electromyograms) is encouraged. In addition, procedures used to decontaminate LFP signals from spiking activity should be described fully.

- Where signals are analyzed by separation into distinct frequency bands, *a priori* reasons for the choice and definition of frequency bands should be provided; alternatively, signals should be analyzed systematically across a wide range of frequencies to avoid biases. Furthermore, it is important to specify whether the observations represent oscillations, band-limited rhythms, broadband spectral changes, or 1/f slopes, and to provide criteria for distinguishing between those inferences, in particular if several of these analyses are performed.

4. Animal state (anesthetized, awake-disengaged, awake-engaged)

Depending on the experimental goals, *in vivo* recordings can be obtained from animals in different states of consciousness, including anesthetized, sleeping, awake, or engaged in a task. In the past, most experiments were performed on anesthetized, head-restrained animals for the purpose of improving single-unit recording stability and duration. However, neurons can display significantly different magnitudes or patterns of discharge as a function of brain state, developmental stage, behavioral task, and experience. Therefore, we suggest the following guidelines for designing *in vivo* physiology experiments and interpreting recorded activity:

- Determine whether a specific relationship between neural activity and behavior is sought, and choose the preparation that is most appropriate for correlating the two measurements.
- When the choice of recording preparation restricts interpretation, an explanation of the limitations should be provided in the Discussion section.
- Describe technical constraints that may favor recordings that have been obtained from animals in a particular state.
- Because neural activity may vary even within a state, quantitative measures of brain state should be provided and compared with the recorded neural activity. For example, proxies for the depth of anesthesia can include breathing and cardiac rate, spectral properties of LFPs, or other EEG-related measures of electrical brain activity. Proxies for motivation during task performance can include trial rate, or false alarm and lapse rates.

5. Data formats that facilitate sharing and meta-analyses

There is increasing, and well-justified, demand for sharing raw data to permit reanalysis, facilitate meta-analysis, and increase transparency. Current barriers to electrophysiological data sharing include multiple, incompatible data formats, some proprietary and many custom-made by individual laboratories; incomplete data formats, which leave out key metadata that are essential to interpret the traces; and a tradition of sharing raw data only for the purpose of a well-defined collaboration.

In the past few years, several efforts have been made toward a comprehensive and open file-format design, with supporting software for reading and writing electrophysiological data. Perhaps the most ambitious effort, Neurodata Without Borders, has been endorsed by multiple research entities, including the Allen Institute for Brain Science, HHMI, and the Simons Foundation. Another approach, including open-source hardware for electrophysiology, as well as data acquisition and analysis software, comes from the Open Ephys project. While there are advantages and dis-

advantages to settling on a single data format, we believe it is important to push toward interoperability, and to promote and plan for data sharing as investigators choose hardware and software for experiments, collect data, and analyze results. Our position is that all researchers should make it easier and more acceptable to share raw data, but we do not mandate a particular format or platform. We urge the community to converge on a set of standards that will make sharing data easier and more common. As always, we invite you to contribute to this discussion by e-mailing *The Journal of Neuroscience* at JN_EiC@sfn.org or tweeting to @MarinaP63.

The Editorial Board of *The Journal of Neuroscience.*
DOI:10.1523/JNEUROSCI.1480–18.2018

References for Further Reading

Dienes Z (2016) How Bayes factors change scientific practice. J Math Psychol 72:78–89. CrossRef

Society for Neuroscience Research Practices for Increasing Scientific Rigor. https://www.sfn.org/Advocacy/Policy-Positions/Research-Practices-for-Scientific-Rigor.

Stead M, Halford JJ (2016) Proposal for a standard format for neurophysiology data recording and exchange. J Clin Neurophysiol 33:403–413. CrossRef Medline

Teeters JL, Godfrey K, Young R, Dang C, Friedsam C, Wark B, Asari H, Peron S, Li N, Peyrache A, Denisov G, Siegle JH, Olsen SR, Martin C, Chun M, Tripathy S, Blanche TJ, Harris K, Buzsáki G, Koch C, et al. (2015) Neurodata without borders: creating a common data format for neurophysiology. Neuron 88:629–634. CrossRef Medline