

Editorial

Considerations for Experimental Design of Behavioral Studies Using Model Organisms

Introduction

Our goal at the *Journal of Neuroscience* is to publish carefully conducted, reproducible studies. To that end, we are publishing a series of editorials on experimental design to contribute to the current discussion on transparency and reproducibility. The field of neuroscience is especially broad; and while there are fundamental principles that guide strong experimental technique across studies, the approaches used to tackle the complexity of the nervous system are particularly diverse, requiring different considerations across subdisciplines. In the current editorial, we address issues related to behavioral experiments in model organisms, such as invertebrates, rodents, birds, and fish. Many issues discussed here are relevant across areas of neuroscience experimentation (e.g., issues of sample size and internal replication) and will likely appear in other editorials as well. We will address issues related to human behavioral studies or those involving nonhuman primates in future editorials because several different considerations apply to these types of studies. Similarly, issues related to *in vivo* electrophysiology or imaging in behaving animals will be considered in a separate editorial.

Reporting

Our goal is for authors to provide sufficient information for other investigators to replicate and extend the study. Like all experiments, the outcome of behavioral studies is influenced by many variables; and the goal of a strong study is to control for those variables as carefully as possible, so full transparency in reporting the conditions under which an experiment is conducted is essential. *Arrive Guidelines* serve as a useful model (<https://www.nc3rs.org.uk/arrive-guidelines>).

For others to reproduce a behavioral experiment, it is first imperative that they are provided with details of the experimental subjects. The species, strain, sex, and age should be noted in the title or abstract. When purchased or acquired from another investigator, the source must be provided because even animals with the same strain name may differ as a result of genetic drift. An RRID (<https://scicrunch.org/resources>) should be included for genetically modified animals (knock-out or transgenic) and newly generated animals should be entered into the RRID database. It is particularly important to describe the breeding strategy used to generate genetically modified organisms. In general, we will not review studies using genetically modified mice without appropriate littermate controls. For invertebrate subjects, the background strain should be carefully described. Finally, for animals that display seasonal variations in behavior (e.g., birds), provide the time of year during which data were collected.

Equally important are husbandry conditions. For example, provide the type of vivarium, whether the facility was specific pathogen free, number of animals per cage or aquarium, and

shape and material of fish tanks. In addition, note the daily temperature range, the light/dark cycle used, and the times of day during which lights were turned on or off.

All details of the study design should be reported, including the definition of the appropriate control groups and what variables were controlled. State the number of subjects in each group, how that number was determined (describe any power analyses), as well as what steps were taken to minimize subjective bias in the observations. For example, note whether individuals were randomly assigned to a particular experimental group and whether the experimenter or behavior scorer was blinded to condition during some or all analyses. With respect to the experimental unit (or the definition of the “*n*” size used), in rodent studies, state whether each animal is considered independently, or whether a litter in a developmental study or a cage in group-housed animals is considered together. For invertebrate studies, note whether an individual or population was considered for each observation. A clear description of software used to gather or analyze behavioral data should be provided, along with the specific measures obtained. After a study is published, custom software must be shared upon request, and we encourage archiving as outlined in our Policy on Computer Code and Software (<http://www.jneurosci.org/content/general-information>).

Most essential for reproducibility is providing comprehensive details of all procedures that were performed. This includes how a study was performed (details of procedures, equipment used, any drug administration or analgesia), when it occurred (time of day when testing was performed and justification of time chosen if animals were tested during their subjective night), and where experiments were conducted (home environment, laboratory, specific apparatus, or testing room). If individuals were tested on more than one task, or on more than one occasion, describe the timeline and history of subjects in each experiment. Information on counterbalancing of treatments, stimuli, and test order between, or within, subjects should be provided. The details of the duration of training or testing, the stimuli used, and interstimulus or intertrial intervals should also be provided. Finally, if any subjects were excluded from analyses, state the number along with the reasons for exclusion.

In reporting of results, we urge authors to show all data points by representation in histograms, scatter plots, or box-and-whisker plots. Authors are encouraged to provide details on effect sizes and confidence intervals. The ability to capture the complexity and variability of phenotypes is a welcome change in the field. It is possible to provide raw numbers for plots as Extended Data, and these data should be in machine readable, widely accessible formats (i.e., CSV, Excel, etc.). Similarly, rather than reporting change from baseline only, relevant baseline data should be reported before testing or treatment.

Statistical design and methods

A full report of statistical design is essential to evaluating the outcome of behavioral studies. Details of the experimental organization (within vs between subjects, longitudinal vs cross-sectional) must be provided, and authors should ensure that the statistical analysis fits the experimental design. For example, results from a study using several experimental groups in a multifactorial design should not be analyzed using one-way ANOVA. Details of the statistical methods used for each analysis should be stated clearly. Indicate between- and within-group factors for ANOVA and any *post hoc* tests, state whether *t* tests were one- or two-tailed, and provide degrees of freedom and test statistic values for each test. In multifactorial designs, statistical interactions that are critical to data interpretation must be provided. When experimental measures are independent of one another, describe the statistical analyses for each separately. However, clearly indicate when many measures are made from the same subject, and describe how the experimental design controls for either multiple comparisons or outcomes that may depend on the experimental timeline. It can be helpful to perform internal replications to validate findings, even if these are reported as a single outcome. If this is done, authors should report the number of replications. We recommend using software for automated checking for errors in statistical reporting (e.g., *statcheck*, <http://statcheck.io>), which will extract all test statistics automatically and determine whether each one of them is internally consistent.

Additional recommendations related to experimental design

We would like to emphasize some recommendations, made above, that are important to unbiased data collection. Whenever possible, it is essential to test blind to condition (genotype, treatment). If this is not possible, describe what steps were taken to

prevent bias, such as use of an automated behavioral apparatus and later blinding of experimenter during analysis. Bias can also be limited by having more than one rater score behavioral outcomes. If this is the case, evaluate and describe methods to ensure interrater reliability. Recording behaviors that are not captured with automated software should be standard so that data are archived (e.g., maintaining videos of the experiment rather than only rating behavior in real time).

In every behavioral study, there are multiple variables that may confound the interpretation of the primary outcome, including, but not limited to, locomotion, arousal, attention, and motivation. It is important to include some evaluation of these variables in the experimental design, and to report how they were controlled. Finally, in complex experimental designs, the behavior itself may confound evaluations performed after testing is completed. Therefore, report the sequence of testing in a complex design, and control for confounds whenever possible.

In conclusion, behavioral studies are integral to many domains of neuroscience. While many of the guidelines described here related to transparent reporting of experimental design and statistical analysis are common across areas of science, there are also many variables that are considered critical for behavioral experiments, such as sex of subjects, time of day, genotype, and history of testing, which are often not considered in other types of studies but can affect the outcome of experiments of many kinds. The overriding goal for investigators, the neuroscience field, and the *Journal of Neuroscience* is for published studies to be replicable, and a full description of design and procedures will help to ensure reproducibility of results.

The Editorial Board of *The Journal of Neuroscience*.
<https://doi.org/10.1523/JNEUROSCI.2794-18.2018>