Received: 6 February 2023 | Revised: 8 May 2023 | Accepted: 27 May 2023

DOI: 10.1111/psyp.14365

ORIGINAL ARTICLE

PSYCHOPHYSIOLOGY SPR WILEY

Managing EEG studies: How to prepare and what to do once data collection has begun

Megan A. Boudewyn[1] | **Molly A. Erickson**[2] | **Kurt Winsler**[3] | **John Daniel Ragland**[4] | **Andrew Yonelinas**[3] | **Michael Frank**[5] | **Steven M. Silverstein**[6] | **Jim Gold**[7] | **Angus W. MacDonald III**[8] | **Cameron S. Carter**[4] | **Deanna M. Barch**[9] | **Steven J. Luck**[3]

[1]Department of Psychology, University of California Santa Cruz, Santa Cruz, California, USA

[2]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, Illinois, USA

[3]Department of Psychology, University of California, Davis, California, USA

[4]Department of Psychiatry and Behavioral Sciences, University of California, Davis, California, USA

[5]Department of Cognitive, Linguistics and Psychological Sciences, Brown University, Providence, Rhode Island, USA

[6]Department of Psychiatry, Neuroscience and Opthamology, University of Rochester, Rochester Medical Center, Rochester, New York, USA

[7]Department of Psychiatry, University of Maryland School of Medicine, Baltimore, Maryland, USA

[8]Department of Psychology, University of Minnesota, Minneapolis, Minnesota, USA

[9]Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, Missouri, USA

Correspondence
Megan A. Boudewyn, Department of Psychology, University of California Santa Cruz, Santa Cruz, CA 95064, USA.
Email: mboudewyn@ucsc.edu

Funding information
National Institute of Mental Health, Grant/Award Number: R01 MH084821, R01 MH084826, R01 MH084828, R01 MH084840, R01 MH084861 and R01 MH087450

Abstract

In this paper, we provide guidance for the organization and implementation of EEG studies. This work was inspired by our experience conducting a large-scale, multi-site study, but many elements could be applied to any EEG project. Section 1 focuses on study activities that take place before data collection begins. Topics covered include: establishing and training study teams, considerations for task design and piloting, setting up equipment and software, development of formal protocol documents, and planning communication strategy with all study team members. Section 2 focuses on what to do once data collection has already begun. Topics covered include: (1) how to effectively monitor and maintain EEG data quality, (2) how to ensure consistent implementation of experimental protocols, and (3) how to develop rigorous preprocessing procedures that are feasible for use in a large-scale study. Links to resources are also provided, including sample protocols, sample equipment and software tracking forms, sample code, and tutorial videos (to access resources, please visit: https://osf.io/wdrj3/).

KEYWORDS

EEG methods, guidelines, large-scale, multisite, protocol, recommendations

M. A. Boudewyn and M. A. Erickson equal contributions to the manuscript.

1 | OVERVIEW

There is a lot to consider when setting up and managing an electroencephalography (EEG) study, beyond the scientific hypothesis being tested. This is particularly true of large-scale and/or multi-site studies, which are becoming increasingly common in the present climate of collaborative science and emphasis on rigor and reproducibility (see Pavlov et al., 2021 for another recent approach). Although there are many obvious advantages to such studies, they inevitably introduce a number of implementation challenges that increase considerably as more participants, experimental tasks, follow-up sessions, data collection sites, and personnel become involved. Alongside a greater opportunity for error, the *cost* of those errors is commensurately larger. Imagine, for instance, that you spend several years planning and conducting a large-scale EEG study, and when data collection is complete you discover that some of your planned analyses cannot be performed because there were errors in the task presentation script, or you discover that the EEG was so noisy that your key effects are not statistically significant despite the large *N*. Similarly, you may find that there are large numbers of artifacts and you must therefore exclude many of your participants from the final analysis, causing your statistical power and generalizability—primary strengths of large studies—to suffer. Alternatively, you may be conducting a smaller-scale study and have a limited amount of time to collect a given number of high-quality datasets (e.g., for a dissertation project). For both large-scale and smaller-scale projects, it is well worth the time and effort to set up robust data collection systems and monitoring procedures to minimize errors and rapidly detect and correct problems when they arise. The set of suggestions that are most useful for a given study will depend on the scale and characteristics of that study.

This paper is motivated by the experience we gained over the course of conducting a large-scale, multi-site EEG study with the Cognitive Neurocomputational Task Reliability And Clinical Applications for Serious Mental Illness (CNTRACS) Consortium. This consortium was convened to identify tasks and computational models that meet the measurement standards necessary for clinical research on cognitive impairment in serious mental illness (Barch et al., 2009; Cohen & Insel, 2008; Luck & Gold, 2008). EEG was collected across five sites from 260 research participants, each of whom completed six experimental tasks, yielding a total of 1560 datasets. Thus, these recommendations focus on preparing for and managing the collection and preprocessing of a large number of datasets, as well as the additional challenges that arise when multiple research teams are involved in data collection. It is worth noting, however, that nearly all of these guidelines are applicable to the organization of any EEG project, regardless of the number of sites or scope of the study. For example, many investigators may find these suggestions useful in building a set of study protocol blueprints to be applied to all studies conducted in their lab, even if each individual experiment may not be considered to be "large-scale." It is also worth noting that these recommendations are our own suggestions rather than official publication guidelines; there are of course other approaches to conducting high-quality EEG studies at any scale (Pavlov et al., 2021). There are also formal guidelines put forth by Society for Psychophysiological Research with respect to EEG analysis and results reporting (Keil et al., 2014, 2022; Picton et al., 2000).

In Section 1, we provide suggestions for establishing the research teams, optimizing the experimental design, setting up the equipment, and training the research staff on EEG data acquisition procedures. In Section 2, our goal is to provide guidance for monitoring data quality and handling a large number of data sets once the study is underway. In particular, we will describe strategies for maintaining consistent and reliable implementation of experimental protocols, monitoring data quality, and developing preprocessing methods that are well suited to a large-scale study. Links to resources are also provided, including sample protocols, sample equipment and software tracking forms, sample code, and tutorial videos (to access resources, please visit: https://osf.io/wdrj3/).

2 | SECTION 1: BEFORE DATA COLLECTION HAS BEGUN

This section covers issues that should be addressed prior to data collection. This includes issues related to team structure, experiment design, equipment and software considerations, training new team members, and communication strategies. A central theme is that your protocol should be designed to minimize the need for even minor deviations over time (or across sites in a multi-site study). In our experience, small exceptions to a protocol compound into major headaches as the scale of the study increases. For example, imagine that the EEG file naming convention is sporadically violated due to unclear instructions, or a subset of participants has an extra EEG recording file without clear documentation about why the extra file exists. Correction of these seemingly minor deviations can require days of effort to track down and account for during data analysis or may produce invalid results. It is more efficient to design and test every element of your study from start to finish before beginning data collection to minimize the number of rules and exceptions that must be hastily constructed on the fly.

2.1 | Before data collection: Establish your teams

Before your study begins, consider how many people will be needed to coordinate the EEG data collection and monitoring activities so that you can budget accordingly. For larger studies, we recommend assembling three teams: a data collection team, a data preprocessing team, and a supervisory team. Data collection team members may also serve on the data preprocessing team; however, these activities require different skills and might be performed at different stages of the study, so we separate them here for the sake of clarity. The responsibilities assigned to each of the three teams are summarized below and covered in more detail Section 2. In a smaller study, you may have a single team (or even a "team of one"), but it is still essential to think carefully about each of these tasks.

2.1.1 | Data collection team

In addition to their primary responsibility of collecting the EEG data, we recommend that all members of the data collection team participate in regular meetings with the EEG supervisory team to review data quality (see Section 2.2: *Conduct Quality Control Assessments*). This team is also responsible for ensuring that (1) all data has been properly backed up at the end of each session, (2) all remarkable events that occurred during the recording session have been documented, and (3) any experiment updates disseminated by the EEG supervisory team have been thoroughly tested.

2.1.2 | Data preprocessing team

After data collection is underway, we suggest training a team of research assistants to do basic EEG preprocessing. Imagine you have 800 datasets (200 participants with 4 tasks each) that take an average of 30 min per dataset to preprocess. This amounts to 20 weeks of effort for someone who spends 50% of their time on EEG preprocessing. Training a team of research assistants to do the preprocessing is a much smaller time commitment, by comparison. Irrespective of the size of the dataset, the data preprocessing team should plan to dedicate a significant amount of their project effort to performing the preprocessing steps (see Section 2.3: *Standardize the Preprocessing Pipeline* for more details).

2.1.3 | EEG supervisory team

Finally, we recommend assembling an EEG supervisory team for a variety of activities ranging from troubleshooting data collection problems to training the team of preprocessors. This team is responsible for the following:

1. For multi-site studies, site visits will be needed to set up equipment (or verify that existing equipment is set up identically across sites), train staff on the data collection protocol, and ensure that everything is working properly (see *Personnel Training*, below). In our experience, an in-person visit by an experienced researcher is ideal to ensure consistency across sites, which in turn saves time later. If this is not possible, workable alternatives may include a combination of remote meetings and video recordings.
2. Someone must be on call during recording sessions to troubleshoot urgent messages from members of the data collection team who are experiencing a recording issue that requires an immediate response (e.g., a broken ground electrode, missing event codes; see *Keeping Lines of Communication Open*). This avoids the need to cancel and reschedule sessions, which is expensive in both time and money.
3. Immediately after the study begins, and any time after there has been a change to the data collection procedures or the task script, an experienced researcher should perform a deep inspection of the first several datasets to ensure that all elements of the task are working as expected. This is especially important for multi-site EEG studies, in which there is ample room for error and miscommunication.
4. Although the goal is to catch and fix any errors before data collection begins, errors may nevertheless occur, or new problems may arise for a number of reasons. Someone will need to dedicate time to fixing such errors for future participants and then developing a method for repairing the data that have already been recorded (see Section 2.4: *Prepare for the Unexpected* for more details).
5. For studies with several principal investigators, you may wish to designate one individual to consolidate information about data quality, number of datasets lost due to artifacts, preliminary results, etc. for investigator meetings. This member of the EEG supervisory team may also provide feedback on data retention directly to the data collection team so that they can make adjustments as needed.
6. Someone will need to oversee implementation and testing of any experiment updates (which may occur if an operating system is updated mid-study, for example). Again, this is a task that becomes much more burdensome with an increasing number of data collection sites, and personnel costs should be budgeted accordingly (see Section 2.4: *Prepare for the Unexpected* for more details).