Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography

ANDREAS KEIL,[a] STEFAN DEBENER,[b] GABRIELE GRATTON,[c] MARKUS JUNGHÖFER,[d]
EMILY S. KAPPENMAN,[e] STEVEN J. LUCK,[e] PHAN LUU,[f] GREGORY A. MILLER,[g] AND CINDY M. YEE[h]

[a]Department of Psychology and Center for the Study of Emotion and Attention, University of Florida, Gainesville, Florida, USA
[b]Department of Psychology, University of Oldenburg, Oldenburg, Germany
[c]Department of Psychology and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois, USA
[d]Institute for Biomagnetism and Biosignalanalysis, University of Münster, Münster, Germany
[e]Center for Mind & Brain and Department of Psychology, University of California, Davis, California, USA
[f]Electrical Geodesics, Inc. and Department of Psychology, University of Oregon, Eugene, Oregon, USA
[g]Department of Psychology, University of California, Los Angeles, California, USA and
Department of Psychology and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois, USA
[h]Department of Psychiatry and Biobehavioral Sciences and Department of Psychology, University of California, Los Angeles, California, USA

Abstract

Electromagnetic data collected using electroencephalography (EEG) and magnetoencephalography (MEG) are of central importance for psychophysiological research. The scope of concepts, methods, and instruments used by EEG/MEG researchers has dramatically increased and is expected to further increase in the future. Building on existing guideline publications, the goal of the present paper is to contribute to the effective documentation and communication of such advances by providing updated guidelines for conducting and reporting EEG/MEG studies. The guidelines also include a checklist of key information recommended for inclusion in research reports on EEG/MEG measures.

Descriptors: Electrophysiology, Methods, Recording techniques, Data analysis, Good practices

Electrophysiological measures derived from the scalp-recorded electroencephalogram (EEG) have provided a window into the function of the living human brain for more than 80 years. More recently, technical advancements allow magnetic fields associated with brain function to be measured as well, resulting in a growing research community using magnetoencephalography (MEG). Although hemodynamic imaging and transcranial magnetic stimulation have expanded the understanding of psychophysiological processes considerably, electromagnetic measures have not lost their importance, largely due to their unparalleled temporal resolution. In fact, recent technological developments have fundamentally widened the scope of methodologies available to researchers interested in electromagnetic brain signals. At the time of writing, the increased use of MEG as well as the development of powerful new hardware and software tools have led to a dramatic growth in the range of research questions addressed. With the growing realization that diverse electrical, magnetic, optical, and hemodynamic psychophysiological methods are complementary rather than competitive, attempts at multimodal neuroimaging integration are growing. In addition, a wide spectrum of data recording, artifact control, and signal processing approaches are currently used, some of which are intimately known only to a small number of researchers. With these richer opportunities comes a growth in demands for technical expertise. For example, there is much greater use of and need for advanced statistical and other signal-processing methods, both adapted from other fields and developed anew. At the same time, the availability of "turn-key" recording systems used by a wide variety of scholars has been accompanied by a trend of reporting less detail about recording and data analysis parameters. Both developments are at odds with the need to communicate experimental procedures, materials, and analytic tools in a way that allows readers to evaluate and replicate the research described in a published manuscript.

The goal of the present paper is to update and expand existing publication guidelines for reporting on studies using measures derived from EEG. If not otherwise specified, these guidelines also apply to MEG measures. This report is the result of the collaborative effort of a committee appointed by the Society for Psychophysiological Research, whose members are listed as the authors of this manuscript. Previous guideline publications and committee reports have laid an excellent foundation for research and publication standards in our field (Donchin et al., 1977; Picton et al., 2000; Pivik et al., 1993). A recent paper by Gross et al. (2013) provided specific recommendations and reporting guidelines for the recording and analysis of MEG data. The reader is referred to these publications in addition to the present report. To

facilitate reading and to assist in educating researchers early in their careers, certain key points will be discussed in the present document, despite having been extensively covered in previous publication guidelines.

How To Use This Document

Publication guidelines and recommendations are not intended to limit the ability of individual researchers to explore novel aspects of electromagnetic data, innovative analyses, or new ways of illustrating results. Rather, the goal of this document is to facilitate communication between authors and readers as well as editors and reviewers, by providing guidelines for how such communication can successfully be implemented. There will be instances in which authors will want to deviate from these guidelines. That will often be acceptable, provided that such deviations are explicitly documented and explained. To aid editors, reviewers, and authors, we have compiled a checklist of key information required when submitting a research report on EEG/MEG measures. The checklist, provided in the Appendix, summarizes central aspects of the guidelines.

Guidelines

Hypotheses and Predictions

In most cases, specific hypotheses and predictions about the electromagnetic activity of interest should be provided in the introduction. Picton et al. (2000) offered helpful examples regarding the presentation of such predictions and their relation to the scientific rationale and behavioral constructs under study. This involves making predictions about exactly how the electromagnetic index will differ by condition, group, or measurement site, reflecting the main hypotheses of the study. It is rarely sufficient to make a general prediction that the measures will differ, without describing the specific ways in which they are expected to differ. For example, a manuscript might describe predicted differences in the amplitude or latency of specific components of the event-related potential (ERP) or event-related field (ERF), or differences between frequency bands contained in EEG/MEG recordings. The predictions should directly relate to the theories and previous findings described in the introduction and describe, to the extent possible, the specific components, time ranges, topography, electrode/sensor sites, frequency bands, connectivity indices, etc., where effects are predicted.

Participant Characteristics

The standards for obtaining informed consent and reporting on human participants, consistent with the Helsinki Accords, the Belmont Report, and the publication manual of the American Psychological Association (APA, 2010), apply fully to electromagnetic studies. It is well established that interindividual differences in the physiological or psychological status of research participants will affect electromagnetic recordings. In fact, such differences are often the focus of a given study. Even in studies not targeting interindividual differences, suitable indicators of participant status should be reported, including age, gender, educational level, and other relevant characteristics. The specifics of what to report may vary somewhat with the nature of the sample, the experimental question, or even the sociocultural context of the research.

In their ERP guidelines report, Picton et al. (2000) discussed areas of importance regarding selection of participants for research and reporting procedures. These include reporting the number of participants, their sensory and cognitive status, and appropriate information on health status. In addition, recent epidemiological work and EEG/MEG research in clinical populations suggest that for certain research questions it is useful to report additional information on the sample. When relying on measures that are sensitive to factors such as psychopathology or alcohol and substance use, for example, additional screening procedures may be appropriate even in studies involving nonclinical samples. Results of epidemiological studies in the United States suggest that past-year and lifetime prevalence rates for a mental disorder are approximately 25% and 46%, respectively (Kessler, Berglund et al., 2005; Kessler, Chiu, Demler, Merikangas, & Walters, 2005). The rate of current illicit drug use in the United States among 18- to 25-year-olds is over 20%, and rates of binge and heavy alcohol use in 21- to 25-year-olds approach 32% and 14%, respectively (SAMHSA, 2012). Among college students, who comprise a significant proportion of research participants in the United States, most estimates place the binge-drinking rate at 40%–50% (e.g., Wechsler & Nelson, 2008). If screening is undertaken, authors should report the method and measures used as well as all inclusion and exclusion criteria.

The issue of matching of groups in clinical studies involving electromagnetic data also warrants careful attention. It has long been noted that matching on one or more characteristics may systematically mismatch groups on other characteristics (e.g., Resnick, 1992). Thus, researchers should carefully consider not only the variables on which to match their groups but the possible unintended consequences of doing so. It may be that a single comparison group cannot handle all relevant issues and that multiple control groups are needed.

Recording Characteristics and Instruments

Evaluation, comparison, and replication of a given psychophysiological study depend heavily on the description of the instrumentation and recording settings used. Many of the relevant parameters are listed in Picton et al. (2000) as well as Gross et al. (2013) for MEG, and the reader is referred to these publications for a discussion of how to report on instrumentation. Here, we recapitulate key information that needs to be provided and offer an update on new requirements and recommendations pertaining to recent developments in hardware and software.

Sensor types. The type of MEG sensor or EEG electrode should be indicated, ideally accompanied by make and model. In addition to traditional sensor types, recent developments in EEG sensor technology include active electrodes. Active electrodes have circuitry at the electrode site that is designed to maintain good signal-to-noise ratio. Electrode-scalp impedances are often of less concern when using active electrodes, and authors may want to emphasize this aspect when using a system that allows reporting impedance values. For passive and active electrodes, the electrode material should be specified (e.g., Ag/AgCl).

Two types of dry sensor technology have become more widely used in recent years (Grozea, Voinescu, & Fazli, 2011). Microspikes penetrate the stratum corneum (Ng et al., 2009), the highly resistive layer of dead cells. Capacitive sensors (Taheri, Knight, & Smith, 1994) rely on conductive materials, such as rubber (Gargiulo et al., 2010), foam, or fabric (Lin et al., 2011). Common to these dry sensor technologies is the challenge of recording the EEG over scalp regions covered by hair. Most previous

applications with dry sensors have involved recordings from the forehead and often have been in the context of research on brain-computer interfaces. When using dry electrodes, the type and technology should be indicated clearly in the manuscript.

Sensor locations. Although MEG sensor locations are fixed relative to each other within a recording system, the position of the participant's head relative to the sensor should be reported along with an index of error/variability of position measurement. When reporting EEG research, electrode positions should be clearly defined. Standard electrode positions include the 10-20 system and the revision to a 10-10 system proposed by the American Electroencephalographic Society (1994). This standard is similar to the 10-10 system of the International Federation of Clinical Neurophysiology (Nuwer et al., 1999). Oostenveld and Praamstra (2001) proposed an extension of the 10-10 system, referred to as the 10-5 system, in order to accommodate electrode arrays with more than 75 electrode sites. An alternative electrode placement system employs a description of the scalp surface based on geodesic (equidistant) partitioning of the head surface with up to 256 positions (e.g., Tucker, 1993), rather than the percentage approach of the 10-20 and related systems. The electrode positions in this system vary according to channel count, as the geodesic partitions differ for different spatial frequencies and ensure regular spacing between electrodes.

Regardless of the system used, a standard nomenclature should be employed and one or more appropriate citations reported. If an equidistant/geodesic placement system is used, the average distance between electrodes should be reported, and the coverage of the head sphere should be described in relation to the 10-20 or 10-10 system. This information can be conveyed in the text or in a figure showing the sensor layout. Ground and reference electrode locations should be specified. For active electrode systems and MEG recordings, the type and location of additional sensors used to reduce ambient and/or subject noise should be indicated. For example, the location of the so-called common mode sense or similar reference electrode should be mentioned if an amplifier system uses such an arrangement. In general, individual reference electrodes are recommended over physically linked reference electrodes (see Miller, Lutzenberger, & Elbert, 1991). Other reference montages can be computed offline.

Spatial sampling. The relationship between the rate of discretization (i.e., digital sampling) and accurate description of the highest frequency of an analog, time-series signal (e.g., the EEG or MEG waveform) is well known as the Nyquist theorem. This theorem posits that signal frequencies equal to or greater than half of the sampling frequency (i.e., the Nyquist frequency) will be misrepresented. This principle of discretization also holds for sampling in the spatial domain for EEG and MEG data, where the signal is the voltage or field data across or above the scalp surface (Srinivasan, Nunez, & Silberstein, 1998). Undersampling in the spatial domain results in high-spatial-frequency features being mistaken for low-spatial frequency information.

In addition to the importance of spatial sampling density, spatial coverage is critical. Traditionally, the head surface is not covered inferior to the axial (horizontal) plane, particularly when using 10-20 EEG positions, whereas equidistant layouts are commonly designed to provide more coverage of the head sphere. In many contexts, especially for source estimation involving much or all of the brain, it is important that the sensor montage extend inferior at least to the equivalent of the axial plane containing F9/F10 in the 10-10 system. At issue is how well local activity from the ventral aspects of the brain is represented. Inadequate spatial sampling can result in a biased estimate of averaged-reference data (Junghöfer, Elbert, Tucker, & Braun, 1999) and misinterpretation regarding the underlying sources (Lantz, Grave de Peralta, Spinelli, Seeck, & Michel, 2003). Authors should address these limitations, particularly when reporting on topographical distributions of electromagnetic data.

Measuring sensor locations. With the increased use of dense sensor arrays and source estimation procedures, the exact location of sensors for a given participant is of increasing importance. Current methods for the determination of 3D sensor positions include manual methods, such as measuring all intersensor distances with digital calipers (De Munck, Vijn, & Spekreijse, 1991), or measuring a subset of sensors arrayed in a known configuration and then interpolating the positions of the remaining sensors (Le, Lu, Pellouchoud, & Gevins, 1998). Another group of methods requires specialized equipment such as electromagnetic digitizers, near-infrared cameras, ultrasound sensors, photographic images acquired from multiple views (Baysal & Sengul, 2010; Russell, Jeffrey Eriksen, Poolman, Luu, & Tucker, 2005), and electrodes containing a magnetic resonance marker together with MR images (e.g., Koessler et al., 2008). When 3D coordinates are reported, the method for obtaining these parameters should be detailed and an index of spatial variability or measurement error provided.

The resolution with which sensor positions can be measured varies considerably, with a potentially strong impact on the validity and reliability of source localization solutions. Under some circumstances, such measurement error contributes most of the source localization error. At best, spatial localization methods are limited largely by the accuracy of sensor position measurement, but compare favorably to the spatial resolution provided by routine functional magnetic resonance imaging (fMRI) procedures (Aine et al., 2012; Miller, Elbert, Sutton, & Heller, 2007). Thus, including the recommended information on measurement of sensor positions is quite important in studies seeking discrete source estimation.

Amplifier type. Accurate characterization of signal amplitude is dependent on many factors, one of which is an amplifier's input impedance. Amplifier systems differing in input impedance vary drastically in their sensitivity to variations in electrode impedance (Ferree, Luu, Russell, & Tucker, 2001; Kappenman & Luck, 2010). Thus, authors are encouraged to report the input impedance of their amplifiers. At minimum, the make and model of the recording system should be indicated in the Methods section.

Impedance levels. If the impedance of the connection between the electrode and the skin is high, this may increase noise in the data. High impedances may also increase the incidence and amplitude of electrodermal (skin) potentials related to sweat gland activity. Traditionally, researchers have addressed these issues by reducing electrode-scalp impedance at each site to be equal to or less than a given threshold (e.g., 10 kΩ). The impedance at a given electrode and time point may be less important than the relationship between electrode impedance and amplifier input impedance in rejecting electrical noise, although sensitivity to skin potentials depends on input impedance even when the amplifier is held constant (Kappenman & Luck, 2010). In addition, the variability and range of impedances across time points and electrodes may add noise to the signal, impacting topographical and temporal information and inferences about sources. EEG systems with active sensors have

different requirements in terms of impedance and may have different ways of reporting indices of data quality. Thus, no single threshold for maximum acceptable electrode impedance can be offered. Reporting should follow the recommendations of the manufacturer. The use of amplifiers with a very high input impedance also reduces the importance of matching impedance for specific sets of electrodes that will be compared, such as at hemispherically homologous sites (Miller et al., 1991).

Obtaining impedances at individual electrodes below a target threshold typically requires preparation (e.g., abrasion) of each site, because the stratum corneum is highly resistive. When such procedures are used, the method should be described. Electrode impedances should be reported when appropriate, or an equivalent signal quality index given when impedances cannot be obtained. When using high-input impedance systems with passive electrodes, information should be given about the range of electrode-scalp impedances, the amplifier's input impedance, and information about how ambient recording conditions were controlled.

Recording settings. The settings of recording devices should be reported in sufficient detail to allow replication. At minimum, parameters should include resolution of the analog-to-digital converter and the sampling rate. In addition, any online filters used during data recording must be specified, including the type of filter and filter roll-off and cut-off values (stated in dB or specifying whether cut-off is the half-power or half-amplitude value; Cook & Miller, 1992).

Stimulus and Timing Parameters

Timing. As noted in Picton et al. (2000), reporting the exact timing of all stimuli and responses occurring during electromagnetic studies is critical. Such information should be provided in a fashion that allows replication of the sequence of events. Required parameters include stimulus durations, stimulus-onset asynchronies, and intertrial intervals, where applicable. Many experimental control platforms synchronize stimulus presentation relative to the vertical retrace of one of the monitors comprising the system, unless otherwise specified. If used, such a linkage should be reported and presentation times accurately indicated in multiples of the video signal retrace (monitor refresh) rate.

When an aspect of the timing is intentionally variable (e.g., variable interstimulus intervals), then a uniform (rectangular) distribution of the time intervals is assumed unless otherwise specified. The relative timing of trials belonging to different experimental conditions should be specified, including any rules or restrictions during randomization, permutation, or balancing. The number of trials in each condition should be explicitly specified, along with the number of trials remaining in each condition when elimination of trials is used to remove artifacts, poor performance, etc. (In many contexts, the number of trials per condition is a primary factor affecting signal-to-noise ratio.) Information about practice trials should also be provided. The total recording duration should be specified along with the duration of rest breaks between recording blocks, where applicable.

Stimulus properties. Replication of a given study is possible only if the relevant parameters defining the stimuli are described. This may include a description of the experimental setting in which the stimuli are presented, including relevant physical and psychosocial aspects of the participants' environment. For instance, it may be appropriate for some studies to report the experimenter's gender, the participants' body posture, or the presence or absence of auxiliary instructions such as to avoid eye blinking during particular time periods of the experiment. Other studies may require precise information regarding ambient lighting conditions, or the size of the recording chamber.

Ideally, examples of the stimuli should be provided in the Methods section, especially when nonstandard or complex stimuli are used. For visual stimuli, parameters may include stimulus size and viewing distance, often together with the visual angles spanned by the stimuli. Because of the strong impact of contrast and intensity of a stimulus on the amplitude and latency of electromagnetic responses, these parameters should be reported where applicable. Different measures of contrast and luminance exist, but often luminance density in units of cd/m^2 can be reported along with either a measure of luminance variability across the experimental display or an explicit measure of contrast, such as the Michelson contrast. Reporting on stimulus color should follow good practices in the respective area of research, which may include specifying CIE coordinates. Additional requirements regarding the display device may exist in studies focusing on color processing. In all instances, the type of display should be reported along with the frame rate or any similar parameter. For example, authors may report that a LED screen of a given make and model with a vertical frame rate of 120 Hz was used. In studies that focus on higher-order processing, it may suffice to specify approximate characteristics of stimulus and display (e.g., "a gray square of moderate luminance," "a red fixation cross").

Similarly, investigators examining responses to auditory stimuli should report the intensity of the stimuli in decibels (dB). Because the dB scale reflects a ratio between two values (not absolute intensity), the report of dB values should include the specification of whether this relates to sound pressure level (SPL), sensation level (SL), hearing level (HL), or another operational definition. When appropriate, the frequency content of the stimulus should be reported along with measures of onset and offset envelope timing and a suitable measure of the energy over time, such as the root mean square. Paralleling the visual domain, the make and model of the delivery device (e.g., headphones, speakers) should be specified. The procedure for intensity calibration should also be reported. If this was adjusted for individual subjects, that procedure should also be reported.

Electromagnetic studies with stimuli in other modalities (e.g., olfactory, tactile) should follow similar principles by defining the nature, timing, and intensity of stimuli and task parameters, per good practice in the respective field of research.

Response parameters. The nature of any response devices should be clearly specified (e.g., serial computer mouse, USB mouse, keyboard, button box, game pad, etc.). Suitable indices of behavioral performance should be indicated, likely including response times, accuracy, and measures of their variability. Because computer operating systems differ in their ability to support accurate response recording, reports may include this information when appropriate.

Data Preprocessing

Data preprocessing commonly refers to a diverse set of procedures that are applied to the data prior to averaging or other major analysis procedures. These procedures may transform the data into a form that is appropriate for more generic computations and eliminate some types of artifact that cannot be dealt with

satisfactorily through averaging methods, spectral analysis, or other procedures. The end product is a set of "clean" continuous or single-trial records ready for subsequent analysis. Reports should include clear descriptions of the methods used for each of the preprocessing steps along with the temporal order in which they were carried out. The following section considers preprocessing steps.

Transformation from A/D units into physical units. This step is necessary for comparing data across experiments. Either the data need to be presented in physical units (e.g., microvolts, femtotesla), or waveforms presented in the paper should report a suitable comparison scale with respect to a physical unit.

Rereferencing. The reference issue is an important problem in EEG studies, and aspects of it were also discussed above (see Recording Characteristics and Instruments). In EEG research, a different reference may be used for online recording and offline analysis. For example, data can be recorded referenced to a specific electrode (e.g., left mastoid), and a different reference (such as average mastoids, or average reference) can be computed offline. Typically, this involves a simple linear transformation corresponding to adding or subtracting a particular waveform from all the channels (unless a Hjorth or Laplacian reference system is used, see Application of Current Source Density or Laplacian Transformations). As data change substantially depending on the reference method used, the type of reference used for online recording and offline analysis should be indicated clearly. When the average of multiple sites is used as the reference, all the sites should be clearly specified, even if they are not used as "active" sites in the analyses.

Interpolation of missing data. Some EEG/MEG channels may contain excessive artifacts and thus may not be usable. For example, an electrode may detach during recording, or a connection may become faulty. The probability of such artifacts increases with dense sensor arrays. It may nevertheless be preferable to include data from missing channels in the analysis; for example, in cases where the analysis software requires identical sensor layouts for all participants. For this reason, problematic or missing data are often replaced with interpolated data. Interpolation is a mathematical technique for estimating unobserved data, according to some defined function (e.g., linear, sphere, spline functions, average of neighbors), from those measured. Interpolation can be used to estimate data between sensor locations (such as those used for color-coded topographic maps) or to replace missing data at a given sensor. For spatial interpolation, it is important to note that the estimated data do not provide higher-density spatial content than what is contained in the original data (Perrin, Pernier, Bertrand, Giard, & Echallier, 1987). Moreover, when interpolation is used to replace missing data, the limit to accuracy is a function of the spatial frequency of the missing data and the number and distribution of sensors. That is, if missing data points are spatially contiguous, and the data to be replaced are predominantly of high spatial content, then it is likely that the estimated data are spatially aliased and do not provide an accurate replacement. Publications should report the interpolation algorithm used for estimating missing channels or for spatial interpolation resulting in topographical maps. Information should be provided as to how many missing channels were interpolated for each participant. Often, information about the spatial distribution of interpolated channels will be required.

Segmentation. Most analysis methods, including traditional cross-trial averages and spectral and time-frequency analyses, are based on EEG/MEG segments of a specific length and a given latency range with respect to an event. Although some data are still recorded in short epochs, data are now more typically recorded in a continuous fashion over extended periods of time and then segmented offline into appropriate epochs. The time range used to segment the data should be reported.

Baseline removal. ERP data are changes in voltage between locations on the recording volume occurring over time. Similarly, ERFs are changes in magnetic field strength. To quantify this change, researchers often define a time period during which the mean activity is to be used as an arbitrary zero value. This period is called the "baseline" period. The mean value recorded during this baseline period is then subtracted or divided from the rest of the segment to result in a measure of change with respect to this zero level. In most ERP/ERF studies, a temporally local baseline value is computed for each recording channel. The choice of baseline period is up to the investigators and should be appropriate to the experimental design. The baseline period should be specified in the manuscript and should ideally be chosen such that it contains no condition-related differences. In addition, as discussed in the section Results Figures, the baseline period should be displayed in waveform plots. Alternative procedures may be used to establish change with respect to a baseline, including regression or filter-based methods. Authors should indicate the method and the data segments used for any baseline procedures. Additional recommendations exist for baseline removal with spectral or time-frequency analyses, as discussed in the Spectral Analysis section.

Artifact rejection. There are many types of artifacts that can contaminate EEG and MEG recordings, including artifacts generated by the subject (e.g., eye blinks, eye movements, muscle activity, and skin potentials) and artifacts induced by the recording equipment or testing environment (e.g., amplifier saturation and line noise). These artifacts are often very large compared to the signal of interest and may differ systematically across conditions or groups of subjects, making it necessary in many experiments to remove the data segments with artifacts from the data to obtain a clean signal for analysis. It should be specified whether artifacts were rejected by visual inspection, automatically based on an algorithm, or a combination of visual inspection and automatic detection. The algorithms used for automatic detection of artifacts should be described in the paper (e.g., a moving window peak-to-peak algorithm). It should also be stated whether thresholds for automatic detection procedures were set separately for individual subjects or channels of data. If any aspect of these procedures was controlled by the experimenter (e.g., manual rejection or subject-specific parameter settings), it should be indicated whether this was done in a manner that was blind to experimental condition or participant group. Unless otherwise specified, it is assumed that all channels are rejected for a segment of data if an artifact is identified in a single channel.

Importantly, because the number of artifacts may differ substantially between experimental conditions or between groups of subjects (e.g., some patient populations may exhibit more artifacts than healthy comparison subjects), the number or percentage of trials rejected for each group of subjects must be specified (especially if peak amplitude measurement is used; see section below on measurement procedures). It should be made clear to the reader whether

the number of trials contributing to the averages after artifact rejection differs substantially across conditions or groups of subjects.

Artifact correction. It is often preferable and feasible to estimate the influence of an artifact on the EEG or MEG signal and to subtract the estimated contribution of the artifact, rather than rejecting the portions of data that contain artifacts. A number of correction methods have been proposed, including regression methods (such as Gratton, Coles, & Donchin, 1983; Miller, Gratton, & Yee, 1988), independent component analysis (ICA; Jung et al., 2000), frequency-domain methods (Gasser, Schuller, & Gasser, 2005), and source-analysis methods (e.g., Berg & Scherg, 1994b). A number of studies have investigated the relative merits of the various procedures (e.g., Croft, Chandler, Barry, Cooper, & Clarke, 2005; Hoffmann & Falkenstein, 2008). By and large, most of these studies have shown that correction methods are generally effective. The major issues are the extent to which they might undercorrect (leaving some of the artifact present in the data) or overcorrect (eliminating some of the nonartifact activity from the data) and whether such correction errors are inconsistent (e.g., larger for sensors near the eyes).

Regardless of which artifact correction procedure is chosen, the paper should provide details of the procedure and all steps used to identify and correct artifacts so that another laboratory can replicate the methods. For example, it is not sufficient to state "artifacts were corrected with ICA." Paralleling artifact rejection (above), other necessary information includes whether the procedure was applied in whole or in part automatically. If nonautomatic correction procedures were used, it should be specified whether the individual performing the procedure was blind to condition, group, or channel. In addition, a detailed list of the preprocessing steps performed before correction (including filtering, rejection of large artifacts, segmentation, etc.) should be described. If a statistical approach such as ICA is used, it is necessary to describe the criteria for determining which components were removed.

If a participant blinks or moves his/her eyes during the presentation of a visual stimulus, the sensory input that reaches the brain is changed, and ocular correction procedures are not able to compensate for the associated change in brain-related processing. In experimental designs where compliance with fixation instructions is crucial (e.g., visual hemifield studies), it is recommended that segments of data that include ocular artifacts during the presentation of the stimuli be rejected prior to ocular correction.

Offline filtering. Filters can be used to improve the signal-to-noise ratio of EEG and MEG data. As filters lead to loss of information, it is often advisable to do minimal online analog filtering and instead use appropriately designed offline digital filters. In general, both online and offline filters work better with temporally extended epochs and therefore may be best applied to continuous than to segmented data. This is particularly the case for high-pass filters used to eliminate very low frequencies (drift) and for low-pass filters with a very sharp roll-off. The type of filters used in the analysis should be reported along with whether they were applied to continuous or segmented data. It is not sufficient to simply indicate the cut-off frequency or frequencies; the filter family and/or algorithm should be indicated together with the filter order and descriptive indices of the frequency response function. For example, a manuscript may report that "a 5th order infinite impulse response (IIR) Butterworth filter was used for low-pass filtering on the continuous (nonsegmented data), with a cut-off frequency (3 dB point) of 40 Hz and 12 dB/octave roll-off." Other ways of reporting filter characteristics are possible, but they should include the filter family (e.g., Boxcar, Butterworth, Elliptic, Chebychev, etc.) as well as information on the roll-off or steepness of the transition in the filter response function (e.g., by indicating that the roll-off was 12 dB/octave). The most common single index of a filter's frequency response function is the half-amplitude or half-power cutoff (the frequency at which the amplitude or power is reduced by 50%). The half-power and half-amplitude frequencies are not the same, so as noted above it is important to indicate whether the cutoff frequency specifies the half-amplitude (–6 dB) point or the half-power (–3 dB) point (Cook & Miller, 1992; Edgar, Stewart, & Miller, 2005).

Measurement Procedures

After preprocessing is completed, the data are typically reduced to a much smaller number of dependent variables to be subjected to statistical analyses. This often consists of measuring the amplitudes or latencies of specific ERP/ERF components or quantifying the power or amplitude within a given time-frequency range. Increasingly, it involves multichannel analysis via principal component analysis (PCA) or ICA, dipole or distributed source analysis, and/or quantification of relationships between channels or sources to evaluate connectivity. Choosing the appropriate measurement technique for quantifying these features is important, and there are a variety of measurement techniques available (some described below; see Fabiani, Gratton, & Federmeier, 2007, Handy, 2005, Kappenman & Luck, 2012b, Kiebel, Tallon-Baudry, & Friston, 2005, or Luck, 2005, for more information). The following guidelines are described primarily in the context of conventional, time-domain ERP/ERF analyses, but many points apply to other approaches as well.

Isolating components. Successful measurement requires care that the ERP/ERF component of interest is isolated from other activity. Even a simple experimental design will elicit multiple components that overlap in time and space, and therefore it is usually important to take steps to avoid multiple components contributing to a single measurement. As a first step, researchers typically choose a time window and a set of channels for measurement that emphasize the component of interest. However, source-analysis procedures have demonstrated that multiple components are active at a given sensor site at almost every time point in the waveform (Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002; Picton et al., 1999). Many investigators have acknowledged this problem and have proposed various methods for addressing it in specific cases (e.g., Luck, 2005). Methods for isolating a single component include creating difference waves or applying a component decomposition technique, such as ICA or spatiotemporal PCA, discussed in the Principal Component Analysis and Independent Component Analysis section (Spencer, Dien, & Donchin, 1999; for general discussion of techniques for isolating components of interest, see Kappenman & Luck, 2012a; Luck, 2005). The method for component isolation used in a particular study should be reported. If no such method is used, authors are encouraged to discuss component overlap as a potential limitation. Most effective in separating the sources of variance, when feasible, is an experimental design that manipulates the overlapping components orthogonally (see Kappenman & Luck, 2012b, for a description).

Description of measurement procedures. The measurement process should be described in detail in the Method section. This

includes specifying the measurement technique (e.g., mean amplitude) as well as the time window and baseline period used for measurement (e.g., "Mean amplitude of the N2pc wave was measured 175 to 225 ms poststimulus, relative to a −200 to 0 ms prestimulus baseline period."). A justification for the time window and baseline period should be included (as described in more detail below).

For measurement techniques that use peak values—such as peak amplitude, peak latency, and adaptive mean amplitude, which relies in part on the peak value—additional information about the measurement procedures is helpful. First, the method used to find the peak value should be specified, including whether the peak values were determined automatically by an algorithm or determined (in whole or in part) by visual inspection of the waveforms. If visual inspection was used, the manuscript should specify whether the individual performing the inspection was blind to condition, group, or channel information. It should also be stated whether the peak was defined as the absolute peak or the local peak (e.g., a point that was greater than the adjacent points even if the adjacent points fell outside the measurement window—see Luck, 2005). It should also be stated whether the peak value was determined separately for each channel or whether the peak latency in one channel was used to determine the measurement latency at the other channels. Generally, it would be misleading to plot the scalp or field distribution of a peak unless it was measured at the same time at all sites.

Peak measures become biased as the signal-to-noise ratio decreases (Clayson, Baldwin, & Larson, 2013), as scores will be more subject to exaggeration due to overlapping noise. In principle, the signal-to-noise ratio improves as a function of the square root of the number of trials in an average. When analyzing cross-trial averages, it is therefore problematic to compare peak values from conditions or groups with significantly different numbers of trials. Studies that rely on peak measures scored from averages (or other measures that may be biased in a similar manner) should report the number of trials in each condition and in each group of subjects. This should include the mean number of trials in each cell, as well as the range of trials included.

A common alternative to peak scoring is area scoring. In general, area measures are less susceptible to signal-to-noise-ratio problems resulting from few (or differing numbers of) trials. An area measure is typically the sum or average of the amplitudes of the measured points in a particular scoring window (after baseline removal). However, the term *area* can be ambiguous, because the area of a geometric shape can never be negative. For measurement techniques that use area measures, it is recommended that the field adopt the following terminology: *positive area* (area of the regions on the positive side of the baseline); *negative area* (area of the regions on the negative side of the baseline); *integrated area* (positive area minus negative area); *geometric area* (positive area plus negative area; this is the same as positive area preceded by rectification).

Inferences about magnitude and timing. Care must be taken in drawing conclusions about the magnitude and timing of the underlying neural activity on the basis of changes in the amplitude and latency of ERPs/ERFs. For example, a change in the relative amplitudes or latencies of two underlying components that overlap in time can cause a shift in the latency of scalp peaks (see Donchin & Heffley, 1978; Kappenman & Luck, 2012a). Thus, a change in peak amplitude does not necessarily imply a change in component magnitude, and a change in peak latency does not necessarily imply a change in component timing. As noted above, suitable methods for

isolating specific spatiotemporal processes assist in the interpretation of overlapping neural events and should be reported in sufficient detail to allow evaluation and replication.

Measurement windows and electrode sites must be well justified and avoid inflation of Type I error rate. EEG/MEG data sets are extremely rich, and it is almost always possible to find differences that are "statistically significant" by choosing measures that take advantage of noise in the data, even if the null hypothesis is true. Opportunities for inflation of Type I error rate (i.e., an increase in false positives) almost always arise when the observed waveforms are used to determine how the data are quantified and analyzed (see Inferential Statistical Analyses). For example, if the time range and sensor sites for measuring a component are chosen on the basis of the timing and topographical distribution of an observed difference between conditions in the same data set, this will increase the likelihood that the difference will be statistically significant even if the difference is a result of noise rather than a reliable effect.

Choosing time windows and sensor sites that maximize sensitivity to real effects while avoiding this kind of bias is a major practical challenge faced by researchers. This problem has become more severe as the typical number of recording sites increased from one to three in the 1970s to dozens at present. Fortunately, conceptual and data-driven approaches have been developed that make it possible to select electrode sites and time windows in a way that is both unbiased and reasonably powerful. An important guideline is to provide an explicit justification for the choice of measurement windows and sensor sites, ensuring that this does not bias results.

A variety of ways of selecting scoring windows and sites are valid. In many cases, the best approach is to use time windows and sensor sites selected on the basis of prior research. When this is not possible, a common alternative approach is to create an average across all participants and conditions and to use this information to identify the time range and topographical distribution of a given component. This is ordinarily an unbiased approach with respect to differences between groups or conditions. However, care must be taken if a given component is larger in one group or condition than in another group or condition or if group or condition numbers differ substantially. For example, Group A may have a larger P3 than Group B, and the average across Group A and Group B is used to select the electrode sites for measuring P3 latency. If Groups A and B differ in P3 scalp distribution, then a suboptimal set of sensor sites would be used to measure P3 latency in Group B, and this could bias the results. The issue of making scoring decisions based on the data being analyzed is considered further in the Inferential Statistical Analyses section.

Another approach is to obtain measures from a broad set of time ranges or sensor locations and include this as a factor in the statistical analysis. For example, one might measure the mean amplitude over every consecutive 100-ms time bin between 200 and 800 ms and then include time as a factor in the statistical analysis. If an effect is observed, then each time window or site can be tested individually, using an appropriate adjustment for multiple comparisons. A recent approach is to use cluster-based analyses that take advantage of the fact that component and time-frequency effects typically extend over multiple consecutive sample points and multiple adjacent sensor sites (see, e.g., Groppe, Urbach, & Kutas, 2011; Maris, 2012).

Both descriptive and inferential statistics should be provided. Reporting of results should include both descriptive and inferential statistics (see Inferential Statistical Analyses). For descriptive

statistics, the group-mean values for each combination of condition and group should be provided, along with a measure of variability (e.g., standard deviation, standard error, 95% confidence interval, etc.). It is generally not sufficient to present inferential statistics that indicate the significance of differences between means without also providing the means themselves. These descriptive statistics can be included in the text, a table, a figure, or a figure caption as appropriate. Descriptive statistics should ordinarily be presented prior to the inferential statistics. This is further discussed in the Inferential Statistical Analyses section.

Results Figures

Electromagnetic data are multidimensional in nature, often including dimensions of sensor channel, topography, or voxel, time and/or frequency band, experimental condition, and group, among others. As a result, particular efforts are necessary to ensure their proper graphical representation on a journal page. Documentation of data quality in most cases will require presentation of a waveform or frequency spectrum in the manuscript. As described in Picton et al. (2000) for ERP data, it is virtually mandatory that figures be provided for the relevant comparisons. They should include captions and labels with all information needed to understand what is being plotted. Ideally, the set of figures in a given paper will graphically convey both the nature of the dependent variable (voltage, field, dipole source strength, frequency, time-frequency, coherence, etc.) and the temporal and spatial properties of the data, as they differ across experimental conditions or groups. As a result, papers will typically have figures highlighting time (e.g., line plots) or spatial distribution (e.g., topographies).

Line plots (see Figure 1) often show change in a measure as a function of time (e.g., amplitude or power of voltages or magnetic fields). Inclusion of line plots is strongly recommended for ERP/ERF studies. It is recommended that waveforms be overlaid, facilitating comparison across conditions, groups, sensors, etc. This may sometimes involve presenting the same waveform twice: for example, patients overlaid with controls for each condition and each condition overlaid separately for patients and controls. Figures should be clearly labeled with the spatial location from which the data were obtained, such as a sensor, source
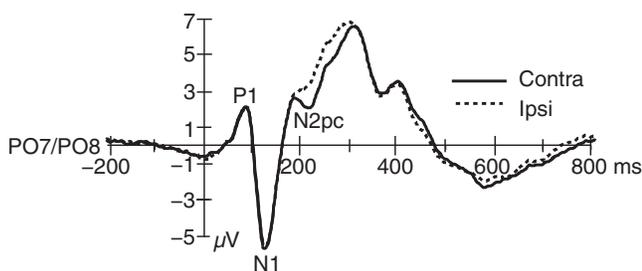
location, or a topographically defined group of sensors. A baseline segment should be included in any data figure that depicts a time course. This baseline segment should be of sufficient length to contain a valid estimate of pre-event or postevent activity. The time segment used as the baseline should be included in the figure (see Data Preprocessing, above). In the case of time-varying power or amplitude in a given frequency band (e.g., event-related (de)synchronization), the baseline segment should contain at least the duration of two cycles of the frequency under consideration (e.g., at least 200 ms of baseline are needed when displaying time-varying 10 Hz activity) or the duration that corresponds to the time resolution of the specific analysis method (e.g., the temporal full width at half maximum of the impulse response).

Each waveform plot should include a labeled x axis with appropriate time units and a labeled y axis with appropriate amplitude units. In most cases, waveforms should include a y-axis line at time zero (on the x axis) and an x-axis line at the zero of the physical unit (e.g., voltage). Generally, unit ticks should be shown, at sufficiently dense intervals, for every set of waveforms, so that the time and amplitude of a given deflection are immediately visible. Line plots displaying EEG segments or ERPs also should clearly indicate polarity, and information regarding both the electrode montage (e.g., 32-channel) and reference (e.g., average mastoid) should be provided in the caption. Including this information in the caption is important to facilitate readers comparing waveforms across studies, which may use different references that contribute to reported differences in the waveforms. The waveforms should be sufficiently large that readers can easily see important differences among them. Where difference waveforms are a standard in a given literature, they may suffice to illustrate a given effect, but in most cases the corresponding condition waveforms should be provided. Because dense arrays of EEG and MEG sensors are increasingly used, it is often not practical to include each sensor's time-varying data in a given figure. For most situations, a smaller number of representative sensors (or sensor clusters) or another suitable figure with temporal information will suffice (e.g., the root mean square). Spatial information related to specific effects may be communicated with additional scalp topographies, as described in the next section.

Scalp potential or magnetic field topographies and source plots are typically used to highlight the spatial distribution of voltage, magnetic fields, current/source densities, and spectral power, among other measures. If the figure includes mapping by means of an interpolation algorithm, such as that suggested by Perrin et al. (1987), the method should be indicated. Interpolation algorithms implemented in some commercial or open-source software suites are optimized for smooth voltage topographies, leading to severe limitations when mapping nonvoltage data or topographies containing higher spatial frequencies (see Data Preprocessing for recommendations on interpolation). Paralleling the recommendations for line plots, captions and labels should accompany topographical figures, indicating the type of data mapped along with a key showing the physical unit. In addition, the perspective should be clearly indicated (e.g., front view, back view; left/right) when using 3D head models. For flat maps, clearly labeling the front/nose and left/right and indicating the projection method in the caption are important. The location of the sensors on a head volume or surface should be visible, to communicate any differences between interpolation across the sensor array, versus extrapolation of data beyond the area covered with sensors. Authors have increasingly used structural brain images overlaid with functional variables

Figure 1. Example of an ERP time series plot suitable for publication. Note that the electrode location is indicated, and both axes are labeled with physical units at appropriate intervals. The onset of the event (in this example, a visual stimulus) is clearly indicated at time zero. Overlaying two experimental conditions fosters comparison of relevant features. (In this example, positive voltage is plotted "up." Both positive up and negative up are common in the EEG/ERP literature, and the present paper makes no recommendation for one over the other.)
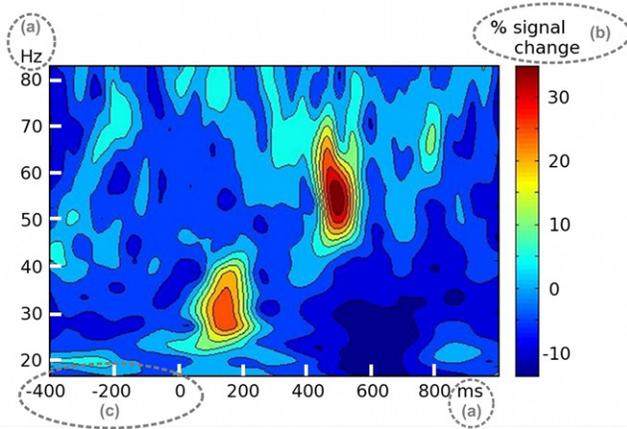
Figure 2. Example of a time-frequency figure suitable for publication. The figure shows a wide range of frequencies (vertical axis) to allow readers a comparison of temporal dynamics in different frequency bands. The vertical axis and the horizontal (time) axis are labeled with appropriate physical units (a), and so is the color bar (b), which here indicates the power change from baseline levels in percent. Hence, it is important that a sufficiently long baseline segment (c) be shown on the plot.

derived from EEG and MEG. In many such cases, a key to relevant anatomical regions (slice location or a reference area) shown should be provided.

Connectivity figures and time-frequency plots are increasingly employed to illustrate different types of time-frequency and connectivity analyses. Although these figures will necessarily vary greatly, inclusion of figures showing original data is recommended, along with the higher-order analyses. Where possible, authors should provide illustrations that maintain a close relationship to the data and analyses performed. For example, results of connectivity analyses carried out for scalp voltages should normally be displayed on a scalp model and not on a standard brain. For time-frequency figures (see Figure 2), a temporal baseline segment should be included of sufficient length to properly display the lowest frequency in the figure, as noted above. In general, many of the recommendations discussed above for line plots and topographical figures apply to other figure types, and authors are encouraged to apply them as appropriate.

Inferential Statistical Analyses

Data analysis is a particularly rich and rapidly advancing area in the EEG/MEG literature. Because electromagnetic data are often multidimensional in nature, statistical analysis may pose considerable computational demands. As noted in Picton et al. (2000), researchers must ensure that statistical analyses are appropriate both to the nature of the data and to the goal of the study. Generally, statistical approaches applied to EEG/MEG data fall into two categories: one where data reduction of dependent variables to relatively small numbers is accomplished on the basis of a priori assumptions, as described above, and one where the number of dependent variables to which inferential statistical testing is applied remains too large to allow meaningful application of traditional statistical methods. Various aspects of these two approaches are discussed below. Approaches in which multivariate methods are used to reduce the dimensionality of the data prior to statistical analysis are discussed in the Principal Component Analysis and Independent Component Analysis section.

Studies with preselected dependent variables. Even when statistical analyses are conducted on a limited number of dependent variables that are defined in advance, authors must ensure the appropriateness of the procedure for the data type that is to be analyzed. Where assumptions of common statistical methods are violated (e.g., normality, sphericity), it should be noted and an adequate correction applied when available. For example, it is expected that the homogeneity/heterogeneity of covariance in within-subjects designs is examined and addressed by Greenhouse-Geisser, Huynh-Feldt, or equivalent correction if the assumption of sphericity is violated (Jennings, 1987) or that multivariate analysis of variance (MANOVA) is undertaken (Vasey & Thayer, 1987). Often, nonparametric statistical approaches are appropriate for a given data type, and authors are encouraged to consider such alternatives. For example, permutation tests and bootstrapping methods have considerable and growing appeal (e.g., Maris, 2012; Wasserman & Bockenholt, 1989). Concerns specific to studies of group effects, as noted in Picton et al. (2000), include the importance of demonstrating that the dependent variable (e.g., an ERP component, a magnetic dipole estimate) is not *qualitatively* different as a function of group or condition. This is an obvious problem when latencies or waveforms of components or spectral events differ by group or condition. In these situations, authors should be careful to ensure that their dependent variables are indeed measuring the same phenomena or constructs across groups, conditions, or brain regions.

Jackknife approaches have been specifically designed to examine latency differences (Miller, Patterson, & Ulrich, 1998), enhancing the signal-to-noise ratio of the measurement by averaging across observations within a group (in $n - 1$ observations) and assessing the variability using all possible $n - 1$ averages. These and similar methods should be considered when latency differences are of interest. Because jackknife methods are based on grand means, they depend greatly on assumptions regarding variability within conditions and groups and on the criterion used to identify the latency of an event (e.g., 50% vs. 90% of an ERP peak value). Thus, authors should report measures of variability as well as indicate the criterion threshold and a rationale for its selection.

In studies designed to evaluate group differences, often as a function of psychopathology or aging, participants may also differ in their ability to perform a task. Task performance by healthy young-adult comparison subjects may be more accurate and faster than that by patients, children, or older participants. Assuming that only electromagnetic data obtained under comparable conditions are to be included in each average (e.g., correct trials only), it is not unusual for psychometric issues concerning group differences in performance to arise. Patient participants, for example, may require more trials (and consequently a longer session) in order to obtain as many correct trials as healthy comparison subjects. Because analyzing sessions of varying durations in a single experiment risks a variety of confounds, identifying and selecting a subset (and equal number) of trials that occur at approximately the same time point for each patient and their demographically matched comparison participant can help to address the issue. Statistical methods that explicitly address group differences in mixed designs, or with nested data in general, include multilevel models (Kristjansson, Kircher, & Webb, 2007). These approaches allow researchers to explicitly test hypotheses on the level of individuals versus groups and separately assess different sources of variability, for instance, regarding their respective predictive value. Researchers reporting on multilevel models should indicate the explicit model used in equation form.

Explicit interactions supporting inferences. Hypotheses or interpretations that involve regional differences in brain activity should be supported by appropriate tests. For example, hemisphere should be included as a factor in statistical analyses, if inferences are made about lateralized effects. In general, such inferences are not justified when the analysis has essentially involved only a simple-effects test for each sensor, voxel, or region of interest. Simple-effects tests may be appropriate for exploring an interaction involving hemisphere but are not themselves a sufficient basis for inferences about lateralization. This example of an effect of hemisphere generalizes to inferences regarding region-specific findings. If two groups or conditions differ in region X but not in region Y, a test for a Group × Region or Condition × Region interaction is usually needed in order to infer that the effect differs in regions X and Y. It may seem logically sufficient to show that the effect in X differs from zero, and the effect in Y does not. In such an analysis, however, it is essential to demonstrate that the confidence intervals do not overlap. Even if the means of X and Y fall on opposite sides of zero, it is possible that their confidence intervals overlap.

Scaling of topographic effects. Because topography is a traditional criterion for defining ERP and ERF components, topographic differences can suggest differences in neural generators. However, even varying amplitudes of a single, invariant neural source can produce Location × Condition interactions in statistical analyses, despite their fundamentally unchanged topography. McCarthy and Wood (1985) brought this issue to the attention of the literature and recommended scaling of the data to avoid the problem, using a normalization of the amplitude of the scalp distribution. This recommendation was widely endorsed for a time. Urbach and Kutas (2002, 2006), however, showed that scaling may eliminate overall amplitude differences between distributions without altering the topography. They argued convincingly that the recommended scaling should not be performed routinely, because it does not resolve the interpretive problem of the source of electromagnetic data. They also demonstrated that standard baseline subtraction further compromises the scaled data.

Although such scaling is no longer recommended, the interpretative problem remains. This issue does not affect inferences that are confined to whether two scalp topographies differ, which may suffice in many experimental contexts. The issue arises, however, when the goal is to make an inference about whether underlying sources differ. Generally, formal source analysis may be needed to address such a question, rather than relying solely on analyses in sensor space.

Handling of baseline levels. Even the routine removal of baseline differences between groups, conditions, and recording sites can potentially be problematic. First, whether to remove baseline variance by subtracting baseline levels or by partialing them out is often not obvious, and the choice is rarely justified explicitly. Second, baseline activity may be related to other activity of interest. If groups, conditions, or brain regions differ at baseline, removal of baseline levels may inadvertently distort or eliminate aspects of the phenomenon of interest (discussed below). It is therefore recommended that any baseline differences between groups or conditions be examined, where appropriate. Third, baseline removal can be particularly problematic for some types of brain source localization analyses. If a source estimation algorithm attempts to identify a source based on the spatial pattern of scalp-recorded signal intensities, but those values have been adjusted by

removal of site-specific baselines, the topography of the scalp signals will in some cases no longer represent the original topography of the source. Urbach and Kutas (2002) discussed this issue in detail as it applies to normalization of topographies, and their critique of baseline removal applies to source analysis as well. Authors are encouraged to assess the robustness of any observed effects against baseline variations and scaling, where appropriate, and to document these steps in the manuscript.

Analytic circularity. A recent controversy in the hemodynamic neuroimaging literature (e.g., Kriegeskorte, Simmons, Bellgowan, & Baker, 2009; Lieberman, Berkman, & Wager, 2009; Viviani, 2010; Vul, Harris, Winkielman, & Pashler, 2009) applies as well to electromagnetic data. When two groups or conditions are compared, and one group or condition is used to define key aspects of the analysis such as the latency window, frequency band, or set of voxels or sensors (topography) for scoring, this approach can bias the results. Resampling methods, replication, and basing scoring decisions on independent data sets can help to avoid this problem. Patterns of results that are contrary to the bias may be accepted as well, providing a particularly conservative test (e.g., Engels et al., 2007).

Psychometric challenges. Electromagnetic measurement is subject to the same psychometric considerations as are other types of measurement, yet issues of task matching, measure reliability, item discriminability, and general versus differential deficit (see Chapman & Chapman, 1973) are often not acknowledged in this literature. These issues go well beyond basic concerns about reliability and validity. For example, given two measurements that differ in reliability yet reflect identical effect sizes, it is generally easier to find a difference with the less noisy measure, suggesting a differential deficit when none may be present. It also is the case that noise in measuring a covariate will tend to lead to underadjustment for the latent variable measured by the covariate (Zinbarg, Suzuki, Uliaszek, & Lewis, 2010). Evaluating the relationship between an independent variable and a dependent variable can be further complicated if a third variable shares variance with either the independent or dependent variable. For example, two groups of subjects may not be ideally matched on age. Even if the discrepancy is not large, the difference may be statistically significant. Age-related changes in electromagnetic activity can occur. Thus, partialing out shared variance provided by the third variable (e.g., age) may not be an option, as the resulting partialed variables may no longer represent the intended construct or phenomenon (Meehl, 1971; Miller & Chapman, 2001). These issues can be particularly difficult to address in clinical or developmental research where random assignment to group is not an option, and groups may differ on variables other than those of interest. Often, explicit modeling of the different effects at different levels may represent a quantitative approach to addressing these issues (Tierney, Gabard-Durnam, Vogel-Farley, Tager-Flusberg, & Nelson, 2012).

Studies with massive statistical testing. Electromagnetic data sets increasingly involve many sensor locations or voxels, time points, or frequency bands and, essentially, massively parallel significance testing. Recent developments in statistical methodology have led to specific procedures to treat such data sets appropriately. The methods may involve calculation of permutation or bootstrapped distributions, without assuming that the data

are uncorrelated or normally distributed (Groppe et al., 2011; Maris, 2012; Mensen & Khatami, 2013). When using such methods, it is recommended that authors report the number of random permutations or a suitable quantitative measure of the distribution used for thresholding, along with the significance threshold employed. Because many variants of massively univariate testing exist, the algorithm generating the reference distribution should be indicated in sufficient detail to allow replication.

Effect sizes. Increasing attention is being paid to effect size and statistical power, beyond the traditional emphasis on significance testing. Effect sizes may be large yet not very useful, or small but theoretically or pragmatically important (Hedges, 2008). The growing attention to effect size, while overdue, can be overdone. First, there is no consensus about what constitutes a "large" or a "small" effect across diverse contexts, although Cohen (1992) provided a commonly cited set of suggestions. Second, effect size is not always important to the research question at hand. In many inferential contexts, the size of the effect is not as important as whether groups or conditions differ reliably, as a basis for inferring whether they are from the same population. Third, there is often little information on which to base a prediction about how large an effect size will be. On the other hand, underpowered studies appear quite commonly, including in the psychophysiology literature. Small-*N* studies are generally limited to finding large effects. Whether that is acceptable in a given case is rarely discussed.

In summary, effect size is often but not always an important issue. When it is important, it should be addressed explicitly. Effect sizes obtained in other contexts rarely suffice to justify an assumption of an effect size anticipated in a given study. Unless a model is employed to make a point prediction of effect size, in most cases what should be discussed is how small an effect a study should be powered to find, not merely what size effect may be likely.

Spectral Analyses

Changes in the spectral properties of raw waveforms during task performance are prima facie properties of EEG and MEG—they are obvious when visually inspecting the ongoing electromagnetic time series or spatial topography. Given its salience even with basic recording setups, the oscillatory character of EEG/MEG and the relationship between different types of oscillations and mental processes have been the focus of pioneering work in EEG (Berger, 1929; in English in 1969) and MEG research (Cohen, 1972). The present discussion will focus on time-domain spectral analyses, but these comments generally apply to spatial spectra as well.

Subsequent to the publication of guidelines for recording and quantitative analysis of EEG (Pivik et al., 1993), the availability of powerful research tools and the increasing awareness that spectral analysis can provide rich information about brain function have led to a sharp increase in the number of scholars interested in spectral properties of electromagnetic phenomena (e.g., Voytek, D'Esposito, Crone, & Knight, 2013). Many recent approaches capitalize on phase information, often used to compute metrics of phase coherence and phase synchrony across trials, time points, or channels. Variables that are reflective of some concept of causality or dependence across spatial location in source or sensor space often are tied to frequency-domain or time-frequency-domain analyses and are therefore discussed in this section as well. Present comments supplement the guidelines presented in Pivik et al. (1993). More recent discussions (Herrmann, Grigutsch, & Busch, 2005; Keil, 2013; Roach & Mathalon, 2008) provide additional resources regarding frequency and time-frequency analyses.

Frequency-domain analyses: Power and amplitude. The frequency spectrum of a sufficiently long data segment can be obtained with a host of different methods, with traditional Fourier-based methods being the most prevalent. These methods are applied to time-domain data (where data points represent a temporal sequence) and transform them into a spectral representation (where data points represent different frequencies), called the frequency domain. Both the power (or amplitude) and phase spectrum of ongoing sine waves that model the original data may be identified in the frequency domain. Most authors use Fourier-based algorithms for sampled, noncontinuous data (discrete Fourier transform, DFT; fast Fourier transform, FFT), the principles of which are explained in Pivik et al (1993; see also Cook & Miller, 1992). Many different implementations of Fourier algorithms exist, defined, for example, by the use of tapering (in which the signal is multiplied with a symmetrical, tapered window function to address distortions of the spectrum due to edge effects), padding (in which zeros, data, or random values are appended to the signal, e.g., to achieve a desired signal duration), and/or windowing (in which the original time series is segmented in overlapping pieces, the spectrum of which is averaged according to specific rules). These steps should be clearly described and the relevant quantities numerically defined. For example, the shape of the tapering window is typically identified by the name of the specific window type. Popular taper windows are Hann(ing), Hamming, Cosine (square), and Tukey. Nonstandard windows should be mathematically described, indicating the duration of the taper window at the beginning and end of the function (i.e., the time it takes until the taper goes from zero to unit level and vice versa). The actual frequency resolution of the final spectrum should be specified in Hz. This is particularly critical when using averaging techniques with overlapping windows, such as the popular Welch-Periodogram method, frequently implemented in commercial software for spectral density estimation. These procedures apply multiple overlapping windows to a time-domain signal and estimate the spectrum for each of these windows, followed by averaging across windows. As a result, the frequency resolution is reduced, because shorter time windows are used, but the signal-to-noise ratio of the spectral estimates is enhanced. Researchers using such a procedure should indicate the type, size, and overlap of the window functions used.

Another important source of variability across studies relates to the quantities extracted from Fourier algorithms and shown in figures and means. Some authors and commercial programs standardize the power or amplitude spectrum by the signal duration or some related variable, to enable power/amplitude comparisons across studies or across different trial types. Other transformations correct for the symmetry of the Fourier spectrum, for example, by multiplying the spectrum by two. All such transformations that affect the scaling of phase or amplitude/power should be reported and a reference to the numerical recipes used should be given. Importantly, the physical unit of the final power or amplitude measure should be given. When using commercial software, it is not sufficient to indicate that the spectrum was calculated using a particular software package; the information described above needs to be provided.

Fourier methods model the data as a sum of sine waves. These sine waves are known as the "basis functions" for that type of analysis. Other basis functions are used in some non-Fourier-based methods, such as nontrigonometric basis functions implemented in

half-wave analysis, for example, where the data serve as the basis function, or in wavelet analysis. In a similar vein, autoregressive modeling of the time series is often used to determine the spectrum. Any such transform should be accompanied by mathematical specification of the basis function(s) or an appropriate reference citation. It is important to note that any waveform, no matter how it was originally generated, can be deconstructed by frequency-domain techniques. The presence of spectral energy at a particular frequency does not indicate that the brain is oscillating at that frequency. It merely reflects the fact that this frequency would be required to reconstruct the time-domain waveform, for example, by combining sine waves or wavelets. Consequently, conclusions about oscillations in a given frequency band should not be drawn simply by transforming the data into the frequency domain and measuring the amplitude in that band of frequencies. Additional evidence is necessary, as discussed in the following paragraphs.

Frequency-domain analyses: Phase and coherence. At any given time, an oscillatory signal is at some "phase" in its cycle, such as crossing zero heading positive. The phase of an oscillatory signal is commonly represented relative to a reference function, typically a sine wave of the same frequency. It is common to assume a reference function that crosses zero heading positive at some reference time, such as the start of an epoch, or stimulus onset, and to describe the difference in phase between the data time series and the reference time series as a "phase lag" of between 0 and 360 degrees or between 0 and 2π radians. Phase and phase lag are also used when two signals from a given data set are compared, whether or not they have the same frequency.

As with spectral power, the phase at a given frequency for a given data segment can be extracted from any time series. When presenting a phase spectrum, authors should detail how the phase information was extracted, following the same steps as indicated above for power/amplitude. Because the reference function is typically periodic, phase cannot be estimated unambiguously. For example, an oscillation may be regarded as lagging behind a half cycle or being advanced a half cycle relative to the reference function, because these two lags would lead to the same time series if the data were periodic. This is typically addressed by so-called unwrapping of the phase. If used, the method for phase unwrapping should be given.

Spectral coherence coefficients can be regarded as correlation indices defined in the frequency domain. They are easily calculated from the Fourier spectra of multiple time series, recorded, for example, at multiple sensors or at different points in time. Correctly reporting the extraction of coherence values may follow the steps for spectral power as described above. Authors using coherence or synchrony measures to address spatial relationships should indicate how they have addressed nonspecific effects such as effects of the reference sensor, volume conduction, or deep dipolar generators, which may lead to spurious coherence between sensors because the generators affect multiple sites (Nolte et al., 2004). Several algorithms have been proposed to address this problem (e.g., Stam, Nolte, & Daffertshofer, 2007) and should be considered by authors. Coherence and intersite synchrony may be better addressed at the source level than at the sensor level, but then information should be included regarding how the source-estimating algorithm addresses the shared-variance problem.

Time-frequency analyses. Spectral analyses as discussed above cannot fully address the issue of overlapping and rapidly changing neural oscillations during behavior and experience. Generally, the

methods above are appropriate only to the extent that the spectral properties of the signal are stable throughout the analyzed interval. Time-frequency (TF) analyses have been developed to handle more dynamic contexts (Tallon-Baudry & Bertrand, 1999). They allow researchers to study changes in the signal spectrum over time. In addition to providing a dynamic view of electrocortical activity across frequency ranges, the TF approach is sensitive to so-called induced oscillations that are reliably initiated by an event but are not consistently phase-locked to it. To achieve this sensitivity, single trials are first transformed into the time-frequency plane and then averaged. TF transformation of the time-domain averaged data (the ERP or ERF) will result in an evolutionary spectrum of the time- and phase-locked activity but will not contain induced oscillations, which were averaged out. As a consequence, the order of averaging steps, if any, should be specified. If efforts are undertaken to eliminate or reduce the effect of the ERP or ERF on the evolutionary spectrum, for example, by subtracting the average from single trials, this should be indicated, and any effects of this procedure should be documented by displaying the noncorrected spectrum. The sensitivity of most TF methods to transient, nonoscillatory processes including ocular artifacts (Yuval-Greenberg, Tomer, Keren, Nelken, & Deouell, 2008) has led to recommendations to analyze a sufficiently wide range of frequencies and display them in illustrations. This recommendation should be followed particularly when making claims about the frequency specificity of a given process.

The Fourier uncertainty principle dictates that time and frequency cannot both be measured at arbitrary accuracy. Rather, there is a trade-off between the two domains. To obtain better resolution in the frequency domain, longer time segments are needed, reducing temporal resolution. Conversely, better time resolution comes at the cost of frequency resolution. Methods developed to address this challenge include spectrograms, complex demodulation, wavelet transforms, the Hilbert transform, and many others.

Providing or citing a mathematical description of the procedure and its temporal and spectral properties is recommended. In most cases, this will be supplied by a mathematical formulation of the basis function (e.g., the sine and cosine waves used for complex demodulation, the parameters of an autoregressive model, or the equation for a Morlet wavelet). Typically, the general procedures are adapted to meet the requirements of a given study by adjusting parameters (e.g., the filter width for complex demodulation; the window length for a spectrogram). These parameters should be given together with their rationale. The time and frequency sensitivity of the resulting function should be specified exactly. For example, the frequency resolution of a wavelet at a given frequency can be indicated by reporting the full width at half maximum (FWHM) of this wavelet in the frequency domain. If multiple frequencies are examined, resulting in an evolutionary spectrum, the relationship between analytic frequencies and their varying FWHM should be reported.

Paralleling frequency-domain analyses, windowing procedures should be specified along with relevant parameters such as the length of the window used. This is particularly relevant for procedures that aim to minimize artifacts at the beginning and end of the EEG/MEG segment to be analyzed. Methods that use digital filters (complex demodulation, event-related desynchronization/event-related synchronization [ERD/ERS], Hilbert transform, etc.) or convolution of a particular basis function (e.g., wavelet analyses) lead to distortions at the edges of the signal, where the empirical time series is not continuous. Even when successfully attenuating

these effects (e.g., by tapering, see above), the validity of these temporal regions for hypothesis testing is compromised, so they should generally not be considered as a dependent variable. Segments that are subject to onset/offset artifacts or tapering should also not be used in a baseline. Therefore, the temporal position of the baseline segment with respect to the onset/offset of the time segment should be indicated. In the same vein, the baseline segment typically cannot be selected to lie in close temporal proximity to the onset of an event of interest, because of the temporal uncertainty of the TF representation, which may change as a function of frequency. Reporting the temporal smearing at the lowest frequency included in the analysis helps readers to assess the appropriateness of a given baseline segment. To facilitate cross-study comparisons, the type of baseline adjustment (subtraction, division, z-transform, dB-transform, etc.) should be specified and the physical unit displayed in figures.

Analyses of phase over time or across recording segments have led to a very productive area of research in which authors are interested in coupling/dependence among phase and amplitude values at the same or different frequencies, within or between brain regions. Many indices exist that quantify such dependencies or interactions in the time domain as well. More recently, inferred causality and dependence approaches have been increasingly used to identify or quantify patterns of connectivity (Greenblatt, Pflieger, & Ossadtchi, 2012). For instance, Granger causality algorithms have been used to index directed dependencies between electromagnetic signals (Keil et al., 2009). Path analysis, structural equation modeling, graph theory, and methods developed for hemodynamic data (e.g., dynamic causal modeling) can readily be applied to electromagnetic time series. Paralleling coherence and synchrony analyses (above), such measures should be described in sufficient detail to allow replication. New algorithms and implementations should be accompanied by extensive, accessible documentation addressing aspects of reliability and validity.

Source-Estimation Procedures

Source-estimation techniques are increasingly used to show that a set of sensor-space (voltage or magnetic field) data is consistent with a given set of intracranial generator locations, strengths, and in some methods orientations. As with other neuroimaging methods, these techniques cannot provide direct and unambiguous evidence about underlying neural activity, such as that the recorded data were actually generated in a given brain area. A wealth of different approaches exists to estimate the intracranial sources underlying the extracranial EEG and MEG recordings. These are known as estimated solutions to the "inverse problem"—inferring the brain sources from remote sensors. (The "forward solution" refers to the calculation of sensor-space data based on given source-space configurations and volume conductor models). The EEG and MEG inverse problem is underdetermined; that is, the generators of measured potentials and fields cannot uniquely be reconstructed without further constraints (Hämäläinen & Ilmoniemi, 1984; von Helmholtz, 1853). Thus, source localization methods provide a *model* of the internal distribution of electrical activity, which should be described. The specific model assumptions, including the constraints chosen by the researcher, should be reported. A mathematical description or a reference to a publication providing such a description should be provided for the steps involved in the source estimation procedure, as outlined below.

Choice and implementation of the head model. All source-estimation approaches rely on a source and a conductivity model of the head. The source model describes the location, orientation, and distribution of possible intracranial neural generators (e.g., gray matter, cerebellum). The conductivity model describes the conductivity configuration within the head that defines the flow of extracellular currents generated by an active brain region, which in combination with intracellular currents eventually leads to measurable scalp potentials or magnetic fields. For any given source and conductivity model, a lead field (i.e., the transfer matrix) can be computed specifying the sensitivity of a given sensor to a given source (e.g., Nolte & Dassios, 2005). The simplest common conductivity model employed in EEG research is a spherical model that describes three to four concentric shells of tissues (brain and/or cerebrospinal fluid [CSF], skull, and scalp), with each tissue volume having homogeneous conductivities in all directions. Because conductivities and thickness of CSF, skull, and scalp do not affect magnetic fields in concentric spherical models, a simple sphere with homogenous conductivity is the easiest model for MEG.

Although simple or concentric spherical models with uniform conductivity assumptions are only rough approximations of realistic conditions, they tend to be feasible and computationally efficient (Berg & Scherg, 1994a). At some cost in computation time and in some cases some cost in anatomic estimates, more realistic models can be generated, with varying resolutions. As more realistic models are employed (e.g., restricting the source model to the gray matter), more constraints are considered in the lead field, thus restricting the number of possible source solutions. An atlas head model, based on geometries of different tissues from population averages, or individual source and conductivity models, based on individual magnetic resonance imaging (MRI)/ computed tomography (CT) data, can be derived. In these cases, a detailed description of the structural images (including acquisition, resolution, segmentation, and normalization parameters) and their transformation into head models is strongly recommended. The potentially high spatial accuracy attainable for individual source/conductivity models should be considered in the context of the limited spatial accuracy affecting other important variables such as sensor position measurements. With less precision, but at far lower cost, the individual head surface can be mapped, for example, with the same equipment that maps sensor locations. To describe the geometries and properties (such as conductivity) of each tissue, boundary element (BEM), finite element (FEM), or finite difference (FDM) methods can be employed, with each model type having specific strengths and weaknesses, which should be discussed in the manuscript (Mosher, Leahy, & Lewis, 1999). Source-estimation software (especially commercially available programs) often provides multiple source and volume conductor models. The conductivity model should be clearly specified and should include (a) the number of tissue types, including their thickness where appropriate; (b) the conductivity (directional or nondirectional) values of each tissue; (c) the method (BEM, FEM, or FDM) if it is a realistic model; and (d) how the sensor positions (such as the average position or positions from individual data) are registered to the head (volume conductor) model. The source model should include the location, orientation, and distribution of the sources.

Description of the source-estimation algorithm. Once source and conductivity models are established, one of two major groups of estimation algorithms is used: focal or distributed source models.

Focal source models (sometimes referred to as dipole models) typically assume that a single or small number of individual dipolar sources—with overall fewer degrees of freedom than the number of EEG/MEG sensors—can account for the observed data (Scherg & Von Cramon, 1986). Estimates from this approach depend on many a priori assumptions (or models) that the user adopts, such as number, rough configuration, and orientation of all active sources, and the segment of data (time course) to be modeled. Once a user establishes a model, a search for the residual source freedoms such as location, orientation, and time course of the source generator(s) is performed, and a solution is established when a cost function is minimized (e.g., the residual variance). A solution can therefore suffer from the existence of local minima. Thus, it is generally advisable that any solution derived from a particular model be tested repeatedly for stability, through repeated runs using the same model parameters but with different starting points. It is recommended that the manuscript describe how stability was assessed. It is important that the model parameters, including the starting conditions and the search method, be specified along with quality indices such as residual variance or goodness-of-fit.

Distributed source methods (in contrast to focal source models) assume that the source locations are numerous and distributed across a given source volume such as the cortical surface or gray matter (Hauk, 2004). The number, spatial density, and orientation of the sources should be specified in the manuscript. The goal is to determine the active sources, with their magnitudes and potentially their orientations. Again, additional assumptions are necessary to select a unique solution. A common assumption is the L1 or L2 minimum norm (least amount of overall source activity or energy, respectively) assumption, but many other assumptions can be employed (such as smoothness in 3D space). Once an assumption is adopted, the user should specify how this assumption is implemented in the solution. Often this step is linked to regularization of the data or the lead field, in which the presence of noise or pragmatic mathematical needs are addressed. The type and strength of regularization should not differ across experimental conditions or groups and should be reported together with an appropriate metric describing its influence on the data.

Spatial filtering is most often implemented as algorithms that involve a variant of so-called beamforming, in which spatial filters are constructed to capture the specific contribution of a given source location to the measured electrical or magnetic field and suppress contributions from other source locations. This approach differs from the algorithms outlined above, as cost functions or fitting the data are not normally involved in beamforming. However, spatial filtering approaches also depend on norms or a priori assumptions. Inherent assumptions of beamforming include a suppression of synchronous activity occurring at different locations. Authors should mention such assumptions and discuss their implications for a given data set. Beamforming depends greatly on the accuracy of the source and conductivity models (Steinstrater, Sillekens, Junghoefer, Burger, & Wolters, 2010), which should be specified in detail, as discussed above. Many types of beamforming exist (Huang et al., 2004), and the exact type used should be mathematically described or a reference to a full description given.

Several points mentioned by Picton et al. (2000) should be briefly highlighted here given their growing importance, as source estimation becomes more common and methods for doing it more varied. Source estimation can be performed on grand-average data, but use on individual participant data is often advisable. Options include using grand-average source estimates as starting points for source estimation done on individual subjects and using different signals for estimating different sources. An explicit rationale should be provided for any such strategy. In many cases, evidence of the reliability (including operator-independence) of the localization should be given, with appropriate figures provided illustrating the variability of the source estimate. The accuracy of a technique may vary across cortical locations (e.g., sulci versus gyri; deep versus superficial sources). Such limitations should be noted, particularly when reporting deep or subcortical sources. Some methods may have desirable properties under ideal conditions, many of which may however not be met in a given study using methodology presently available. To document the validity of a given source estimation approach, it is therefore not sufficient to refer to previous studies of the accuracy of a given technique unless the conditions of the present study are comparable (e.g., signal-to-noise ratio, number and position of sensors, availability of single-subject structural MRI scans).

Principal Component Analysis and Independent Component Analysis

PCA. Principal component analysis has a long history in EEG/MEG research. Its recommended use is described in detail in Dien (2010), Donchin & Heffley (1978), and Picton et al. (2000). Spatial, temporal, and combined variants are often used (Tenke & Kayser, 2005). The structure and preprocessing of the EEG/MEG data submitted to PCA should be described. Comments in the next section about preprocessing for ICA apply for the most part to PCA as well. The specific PCA algorithm used should be described, including the type of association matrix as well as whether and how rotation was applied. Initial PCA and subsequent rotation are separate steps, and both should be described (e.g., "PCA followed by varimax rotation were employed"). In addition, the decision rule for retaining versus discarding PCA components should be provided.

ICA. Independent component analysis refers not to a specific analysis method but a family of linear decomposition algorithms. It has been brought to the EEG/MEG literature much more recently than PCA and is increasingly used for biosignal analysis. Assuming a linear superposition of (an unknown number of) signals originating from brain and nonbrain sources, the aim of ICA in multichannel EEG/MEG data decomposition is typically to disentangle these source contributions into maximally independent statistical components (Makeig, Jung, Bell, Ghahremani, & Sejnowski, 1997; Onton, Westerfield, Townsend, & Makeig, 2006). ICA is now frequently used to attenuate artifacts in EEG/MEG recordings (see artifact correction section above) and, more recently, to identify and distinguish signals from different brain sources. Compared to PCA, ICA solutions do not require a rotational postprocessing step to facilitate interpretation, as independence imposes a more severe restriction than orthogonality (the latter requires second-order statistical moments, the former includes higher-order moments). For the same reason, however, ICA decompositions can be computationally demanding. Introductions to ICA can be found in Hyvärinen, Karhunen, and Oja (2001) and Onton et al. (2006).

Various ICA algorithms and implementations are available. They have been reported to produce generally similar results, but they estimate independence in different ways. The use of different algorithms may also entail different practical considerations.

Among the most popular ICA algorithms for EEG and MEG analysis are infomax ICA (Bell & Sejnowski, 1997) and fastICA (Hyvärinen, 1999). Several implementations exist in different commercial and open source software packages. Since default parameters may vary between implementations and algorithms, the algorithm and software implementation should be reported in the manuscript. Moreover, all parameter settings should be reported. Some iterative ICA algorithms may not converge to the same solution after repeated decomposition of the same data. It is therefore necessary to confirm the reliability of the ICA components obtained. Dedicated procedures have been developed to do so. They should be applied and described (Groppe et al., 2009).

EEG/MEG data are typically of high dimensionality and can be arranged in various different ways before being submitted to ICA. As with PCA, the structure and preprocessing of the EEG/MEG data submitted to ICA should be detailed in the manuscript. The most common data arrangement is a 2D Channels × Time Points matrix structure. Group, subject, condition, etc., may be higher-order dimensions within which the channel dimension is nested. Using ICA this way, the aim is to achieve maximally temporally independent time series. Regarding the first dimension (channels) in such an approach, it is usually good practice to use all recorded EEG channels (returning good-quality voltage fluctuations) for the decomposition. If some channels are excluded from ICA training, this should be reported.

Although dense-array EEG/MEG recordings provide higher-dimensional data and thus in principle enable a more fine-grained decomposition, higher-dimensionality decomposition is computationally more demanding. If the dimensionality of dense-array EEG\MEG data is reduced before ICA decomposition, this should be specified, and the model selection justified. Good results for the separation of some basic EEG features have been achieved for low-density EEG recordings as well. The second dimension submitted to the decomposition, that is, the time points, may be for example the raw voltage time series as originally recorded. It could also represent the preprocessed (e.g., high-pass filtered, rereferenced) raw data, a subset of the raw data after the removal of particular artifacts (e.g., severe, nonrepetitive artifacts), or a subset of the raw data where a particular brain activity pattern is expected to dominate (e.g., the intervals following some events of interest). For most algorithms it is necessary that the second dimension be significantly larger than the first to achieve a satisfactory decomposition result and a reliable solution (Onton et al., 2006). Accordingly, which and how many data points are selected for ICA should be described.

ICA decomposition quality depends strongly on how well the data comply with the statistical assumptions of the ICA approach. How ICA decomposition quality can be evaluated is a matter of ongoing discussion (Delorme, Palmer, Onton, Oostenveld, & Makeig, 2012; Groppe et al., 2009; Onton et al., 2006). One statistical assumption of the ICA approach is the covariance stationarity of the data, which is for example violated by very low-frequency artifacts such as drift. Accordingly, the application of a high-pass filter, de-trending, or de-meaning of the input data is recommended to improve ICA reliability and decomposition quality. These steps should be described. A key feature for the evaluation of decomposition quality is the dipolarity of the inverse weights characterizing the independent components. Infomax ICA appears to outperform other algorithms in this respect (Delorme et al., 2012), and dipolar ICA components have been found to be more reliable than nondipolar ones (Debener, Thorne, Schneider, & Viola, 2010). Given that dipolar projections are biophysically plausible assumptions for spatially circumscribed brain generators (Onton et al., 2006), these two criteria, component dipolarity and component reliability, should be carefully considered in ICA outcome evaluation.

The main ICA outcome, the unmixing weights, may be regarded as a set of spatial filters, which can be used to identify one or several statistical sources of interest. It is important to recognize that sign and magnitude of the raw data are arbitrarily distributed between the resulting temporally maximally independent component activations and the corresponding spatial projections (inverse weights). Accordingly, the grouping of inverse weights across subjects and the grouping of component activations (or component activation ERPs or ERFs) across subjects into a group average representation require some form of normalization and definition of polarity. Alternatively, a back-projection of components of interest to the sensor level solves the sign and magnitude ambiguity problem and produces signals in original physical units and polarity (e.g., microvolts or femtotesla). If taken, these procedural steps should be documented in the manuscript.

The above-mentioned approach requires that the single-subject or single-session EEG/MEG data be submitted to ICA. Accordingly, the decomposition results will differ by subject to some extent, requiring some form of clustering to identify which components reflect the same process. How the component selection process was guided should be specified. For some biological artifacts such as eye blinks, eye movements, or electrocardiac activity, this grouping or clustering process can be achieved efficiently and in an objective procedure by the use of templates (Viola et al., 2009). Here, the inverse unmixing weights of independent components reflect the projection strength of a source to each EEG/MEG channel and thus can be plotted as a map. The spatial properties of these maps can guide component interpretation, as they should show some similarity to the EEG/MEG features a user might be interested in. The identification of ICA components reflecting brain signals should be based not only on spatial information (inverse weights) but also on temporal information (ICA activations). How much (spatial, temporal, or a combination of both) variance or power in the raw signal or the time- or frequency-domain averaged signal a component explains can be determined and can guide the ICA component selection step. All thresholds used for selection should be indicated. Given that the assumption of an equal number of sensors and sources is likely violated in ICA decompositions of real EEG/MEG data, authors should specify how many components representing a given brain activity pattern were considered per subject.

Multimodal Imaging/Joint Recording Technologies

Diverse measures of brain function, including noninvasive neuroimaging such as EEG, MEG, and fMRI, have different strengths and weaknesses and are often best used as complements. EEG and MEG can readily be recorded simultaneously. Increasingly, EEG and MEG analysis is being augmented with per-subject structural MRI recorded in separate sessions. The concurrent recording of electromagnetic and hemodynamic measures of brain activity is a rapidly evolving field. Electromagnetic and hemodynamic measures are also sometimes recorded from the same subjects performing the same tasks in separate sessions.

Different imaging modalities reflect different aspects of brain activity (e.g., Nunez & Silberstein, 2000), which makes their integration difficult, although the fact that they are substantially

nonredundant also makes their integration appealing. As described above for single-modality data, hypotheses must be clearly stated and all preprocessing steps fully described for each measure used. A number of specific issues should also be considered in multimodal imaging studies. Importantly, practical issues of subject safety apply (Lemieux, Allen, Franconi, Symms, & Fish, 1997; Mullinger & Bowtell, 2011). Therefore, only certified hardware should be used for EEG recordings in an MRI system, and the hardware should be specified in the manuscript.

Artifact handling should be detailed. EEG data recorded during MRI are severely contaminated by specific types of artifacts, not covered above (Data Preprocessing), such as the cardioballistic artifact. How these artifacts are handled determines signal quality. The processing of MRI-specific and MRI-nonspecific artifacts should be described. This includes a description of the software used, the order of the signal processing steps taken, and the parameter settings applied.

Single-modality results should be reported. Multimodal integration is a quickly evolving field, and new analysis procedures are being developed that aim to optimize the integration or fusion of information from different modalities (Huster, Debener, Eichele, & Herrmann, 2012). The underlying rationale is that to some extent different modalities capture different aspects of the same or related neural processes. Given that the amount of overlap between modalities is to a large extent unknown, and given that signal quality from multimodal recordings may be compromised, it is generally necessary to report single-modality findings in addition to results reporting multimodality integration. For EEG-fMRI integration, for example, EEG or ERP-alone findings should typically be described, illustrated, and statistically analyzed in addition to the multimodal integration analysis, which might be based on the same (e.g., ERP) or related (single-trial EEG) signals.

Application of Current Source Density or Laplacian Transformations

As discussed in Source-Estimation Procedures, for any given distribution of EEG/MEG data recorded at the surface of the head there exists an infinite number of possible source configurations inside the head. Another theorem, however, maintains that the potential of magnetic field distribution on any closed surface, which encloses all the generators, can be determined uniquely from the extracranial potential or magnetic field map. In MEG, this procedure can, for example, be applied to project the magnetic field distribution of a subject-specific sensor configuration onto a standard configuration. Because there are no neural sources between the scalp and the cortical surface, mapping procedures can be applied to EEG data to estimate the potential distribution on the cortical surface (Junghöfer, Elbert, Leiderer, Berg, & Rockstroh, 1997). This mathematical transformation called "cortical mapping" (CM) can compensate for the strong blurring of the electrical potential distribution, which is primarily a consequence of the low conductivity of the skull. Magnetic fields, in contrast, are almost unaffected by conductivity properties of intervening tissues. MEG topographies thus reveal higher spatial frequencies, convergent to the CM topography, and sensor, scalp, and cortex MEG topographies do not differ much. The uniqueness of CM does not depend on adequate modeling of conductivities or the head shape. Inadequate modeling will, however, give rise to an inaccurate estima-

tion of the cortical surface potential. Authors should therefore describe the model parameters (e.g., assumed conductivities) in the manuscript.

An alternative to cortical mapping that also compensates for the spatial low-pass filter effect reflecting the signal transition between cortex and scalp in EEG is the current source density (CSD) calculation that is based on the spatial Laplacian or the second spatial derivative of the scalp potential (Tenke & Kayser, 2005). Because the reference potential, against which all other potentials are measured, is extracted by the calculation of a spatial gradient—that is, constant addends are removed—both CM and CSD measures are independent of the reference choice. They constitute "reference-free" methods—convergent with reference-free MEG. As a result, CM and CSD may serve as a bridge between reference-dependent scalp potentials and the estimation of the underlying neural generators. Many implementations exist, and authors should indicate the software and parameter settings used to calculate CM and CSD maps.

Importantly, CM and CSD transformations of EEG data and also projections of MEG topographies are unique and accurate if and only if the electrical potential or magnetic field is known for the entire surface surrounding all neural generators. The signal, however, can only be measured at discrete locations covering much less than the entire head. Data at other locations are then estimated through interpolation, in order to compute the CM or the CSD. In the context of CSD mapping, the most common interpolation functions are three- or two-dimensional spline functions, which meet the physical constraint of minimal energy. Moreover, in contrast to the method of nearest neighbors, which has often been used in the past, these functions take the complete distribution at all sensors into consideration. There exist a large set of such functions that all meet this criterion and interpolate the topography in different ways, resulting in different projection topographies. Thus, authors should indicate the interpolation functions that were used. In the same vein, the CM or CSD algorithms applied in a given study should be reported, for example, by providing a brief mathematical description together with a reference to the full algorithm or by describing the full mathematical formulation.

Interpolation-dependent effects are smaller with less interpolation, that is, with denser sensor configurations (Junghöfer et al., 1997). Since CM and CSD act as spatial high-pass filters for EEG topographies, these projections are quite vulnerable to high-spatial-frequency noise (e.g., noise with differential effects on neighboring electrodes). Even if dense-array electrode configurations (e.g., 256-sensor systems) may spatially oversample the brain-generated aspect of the scalp topography, such configurations may reduce high-spatial-frequency noise as a consequence of spatial averaging. Thus, although CM and CSD procedures can be applied with sparse electrode coverage, estimates of CM and CSD measures gain stability with increasing electrode density. Accordingly, authors are encouraged to address the specific vulnerability of CM and CSD to high-spatial-frequency noise with respect to the specific electrode configuration used in a given study.

Single-Trial Analyses

Given recent advances in recording hardware and signal processing as described in the previous sections, researchers have increasingly capitalized on information contained in single trials of electromagnetic recordings. Many such applications exist now, ranging from mapping or plotting routines that graphically illustrate properties of single trials (voltage or field, spatial distribution, spectral phase or

power) to elaborate algorithms for extracting specific dependent variables (e.g., component latency and amplitude) from single-trial data. Recent advances in statistics such as multilevel modeling have also enabled analyses of electromagnetic data on the single trial level (Zayas, Greenwald, & Osterhout, 2010).

Because single-trial EEG and MEG typically have low signal-to-noise ratios, authors analyzing such data are encouraged to address the validity and reliability of the specific procedures employed. The type of temporal and/or spatial filters used is critical for the reliability and validity of latency and amplitude estimates of single-trial activity (e.g., De Vos, Thorne, Yovel, & Debener, 2012; Gratton, Kramer, Coles, & Donchin, 1989). Thus, preprocessing steps should be documented in particular detail when single-trial analyses are attempted. Where multivariate or regression-based methods are used to extract tendencies inherent in single-trial data, measures of variability should be given. Where feasible, readers should be enabled to relate the extracted data to an index with greater signal-to-noise ratio such as a time-domain average (ERP or ERF) or a frequency spectrum. To foster replication, a research report should include a mathematical description of the algorithm employed to extract single-trial parameters or a reference to a paper where such a description is provided.

Conclusion

As discussed in the introductory section, the present guidelines are not intended to discourage the application of diverse and innovative methodologies. The authors anticipate that the spectrum of methods available to researchers will increase rapidly in the future. It is hoped that this document contributes to the effective documentation and communication of such methodological advances.

References

Aine, C. J., Sanfratello, L., Ranken, D., Best, E., MacArthur, J. A., & Wallace, T. (2012). MEG-SIM: A web portal for testing MEG analysis methods using realistic simulated and empirical data. *Neuroinformatics*, *10*, 141–158. doi: 10.1007/s12021-011-9132-z

American Electroencephalographic Society (1994). American Electroencephalographic Society. Guideline thirteen: Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *11*, 111–113.

American Psychological Association (APA) (2010). *The publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Baysal, U., & Sengul, G. (2010). Single camera photogrammetry system for EEG electrode identification and localization. *Annals of Biomedical Engineering*, *38*, 1539–1547. doi: 10.1007/s10439-010-9950-4

Bell, A. J., & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research*, *37*, 3327–3338. doi: 10.1016/S0042-6989(97)00121-1

Berg, P., & Scherg, M. (1994a). A fast method for forward computation of multiple-shell spherical head models. *Electroencephalography and Clinical Neurophysiology*, *90*, 58–64.

Berg, P., & Scherg, M. (1994b). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, *90*, 229–241.

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen [On the electroencephalogram of man]. *Archiv für Psychiatrie und Nervenkrankheiten*, *87*, 527–570.

Berger, H. (1969). On the electroencephalogram of man. [Supplement]. *Electroencephalography and Clinical Neurophysiology*, *28*, 37–73.

Chapman, L. J., & Chapman, J. P. (1973). Problems in the measurement of cognitive deficit. *Psychological Bulletin*, *79*, 380–385.

Clayson, P. E., Baldwin, S. A., & Larson, M. J. (2013). How does noise affect amplitude and latency measurement of event-related potentials (ERPs)? A methodological critique and simulation study. *Psychophysiology*, *50*, 174–186. doi: 10.1111/psyp.12001

Cohen, D. (1972). Magnetoencephalography: Detection of the brain's electrical activity with a superconducting magnetometer. *Science*, *175*, 664–666.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Cook, E. W., III, & Miller, G. A. (1992). Digital filtering: Background and tutorial for psychophysiologists. *Psychophysiology*, *29*, 350–367.

Croft, R. J., Chandler, J. S., Barry, R. J., Cooper, N. R., & Clarke, A. R. (2005). EOG correction: A comparison of four methods. *Psychophysiology*, *42*, 16–24. doi: 10.1111/j.1468-8986.2005.00264.x

De Munck, J. C., Vijn, P. C., & Spekreijse, H. (1991). A practical method for determining electrode positions on the head. *Electroencephalography and Clinical Neurophysiology*, *78*, 85–87.

De Vos, M., Thorne, J. D., Yovel, G., & Debener, S. (2012). Let's face it, from trial to trial: Comparing procedures for N170 single-trial estimation. *Neuroimage*, *63*, 1196–1202. doi: 10.1016/j.neuroimage.2012.07.055

Debener, S., Thorne, J., Schneider, T. R., & Viola, F. C. (2010). Using ICA for the analysis of EEG data. In M. Ullsperger & S. Debener (Eds.), *Integration of EEG and fMRI*. New York, NY: Oxford University Press.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS One*, *7*, e30135. doi: 10.1371/journal.pone.0030135PONE-D-11-14790

Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping*, *15*, 95–111. doi: 10.1002/hbm.10010

Dien, J. (2010). Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations. *Psychophysiology*, *47*, 170–183. doi: 10.1111/j.1469-8986.2009.00885.x

Donchin, E., Callaway, E., Cooper, R., Desmedt, J. E., Goff, W. R., & Hillyard, S. A. (1977). Publication criteria for studies of evoked potentials in man. In J. E. Desmedt (Ed.), *Attention, voluntary contraction and event-related cerebral potentials* (Vol. 1, pp. 1–11). Brussels, Belgium: Karger.

Donchin, E., & Heffley, E. F. (1978). Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related brain potentials research* (pp. 555–572). Washington, DC: U.S. Environmental Protection Agency.

Edgar, J. C., Stewart, J. L., & Miller, G. A. (2005). Digital filtering in EEG/ERP research. In T. C. Handy (Ed.), *Event-related potentials: A handbook*. Cambridge, MA: MIT Press.

Engels, A. S., Heller, W., Mohanty, A., Herrington, J. D., Banich, M. T., & Webb, A. G. (2007). Specificity of regional brain activity in anxiety types during emotion processing. *Psychophysiology*, *44*, 352–363. doi: 10.1111/j.1469-8986.2007.00518.x

Fabiani, M., Gratton, G., & Federmeier, K. D. (2007). Event-related brain potentials: Methods, theory, and applications. In J. T. Cacioppo & L. G. Tassinary (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 85–119). New York, NY: Cambridge University Press.

Ferree, T. C., Luu, P., Russell, G. S., & Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clinical Neurophysiology*, *112*, 536–544. doi: 10.1016/S1388-2457(00)00533-2

Gargiulo, G., Calvo, R. A., Bifulco, P., Cesarelli, M., Jin, C., & Mohamed, A. (2010). A new EEG recording system for passive dry electrodes. *Clinical Neurophysiology*, *121*, 686–693. doi: 10.1016/j.clinph.2009.12.025

Gasser, T., Schuller, J. C., & Gasser, U. S. (2005). Correction of muscle artefacts in the EEG power spectrum. *Clinical Neurophysiology*, *116*, 2044–2050. doi: 10.1016/j.clinph.2005.06.002

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484.

Gratton, G., Kramer, A. F., Coles, M. G., & Donchin, E. (1989). Simulation studies of latency measures of components of the event-related brain potential. *Psychophysiology*, *26*, 233–248.

Greenblatt, R. E., Pflieger, M. E., & Ossadtchi, A. E. (2012). Connectivity measures applied to human brain electrophysiological data. *Journal of Neuroscience Methods*, *207*, 1–16. doi: 10.1016/j.jneumeth.2012.02.025

Groppe, D. M., Makeig, S., & Kutas, M. (2009). Identifying reliable independent components via split-half comparisons. *Neuroimage*, *45*, 1199–1211. doi: 10.1016/j.neuroimage.2008.12.038

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*, 1711–1725. doi: 10.1111/j.1469-8986.2011.01273.x

Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., & Jensen, O. (2013). Good practice for conducting and reporting MEG research. *Neuroimage*, *65*, 349–363. doi: 10.1016/j.neuroimage.2012.10.001

Grozea, C., Voinescu, C. D., & Fazli, S. (2011). Bristle-sensors—low-cost flexible passive dry EEG electrodes for neurofeedback and BCI applications. *Journal of Neural Engineering*, *8*, 025008. doi: S1741-2560(11)76848-3

Hämäläinen, M., & Ilmoniemi, R. (1984). *Interpreting measured magnetic fields of the brain: Estimates of of current distributions*. Helsinki, Finland: Helsinki University of Technology.

Handy, T. C. (2005). *Event-related potentials: A handbook*. Cambridge, MA: MIT Press.

Hauk, O. (2004). Keep it simple: A case for using classical minimum norm estimation in the analysis of EEG and MEG data. *Neuroimage*, *21*, 1612–1621. doi: 10.1016/j.neuroimage.2003.12.018

Hedges, L. V. (2008). What are effect sizes and why do we need them? *Child Development Perspectives*, *2*, 167–171. doi: 10.1111/J.1750-8606.2008.00060.X

Herrmann, C. S., Grigutsch, M., & Busch, N. A. (2005). EEG oscillations and wavelet analysis. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 229–259). Cambridge, MA: MIT Press.

Hoffmann, S., & Falkenstein, M. (2008). The correction of eye blink artefacts in the EEG: A comparison of two prominent methods. *PLoS One*, *3*, e3004. doi: 10.1371/journal.pone.0003004

Huang, M. X., Shih, J. J., Lee, R. R., Harrington, D. L., Thoma, R. J., & Weisend, M. P. (2004). Commonalities and differences among vectorized beamformers in electromagnetic source imaging. *Brain Topography*, *16*, 139–158.

Huster, R. J., Debener, S., Eichele, T., & Herrmann, C. S. (2012). Methods for simultaneous EEG-fMRI: An introductory review. *Journal of Neuroscience*, *32*, 6053–6060. doi: 10.1523/JNEUROSCI.0447-12.2012

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, *10*, 626–634. doi: 10.1109/72.761722

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. New York, NY: John Wiley & Sons.

Jennings, J. R. (1987). Editorial policy on analyses of variance with repeated measures. *Psychophysiology*, *24*, 474–475.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., & Iragui, V. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*, 163–178.

Junghöfer, M., Elbert, T., Leiderer, P., Berg, P., & Rockstroh, B. (1997). Mapping EEG-potentials on the surface of the brain: A strategy for uncovering cortical sources. *Brain Topography*, *9*, 203–217.

Junghöfer, M., Elbert, T., Tucker, D. M., & Braun, C. (1999). The polar average reference effect: A bias in estimating the head surface integral in EEG recording. *Clinical Neurophysiology*, *110*, 1149–1155.

Kappenman, E. S., & Luck, S. J. (2010). The effects of electrode impedance on data quality and statistical significance in ERP recordings. *Psychophysiology*, *47*, 888–904. doi: 10.1111/j.1469-8986.2010.01009.x

Kappenman, E. S., & Luck, S. J. (2012a). ERP components: The ups and downs of brainwave recordings. In S. J. Luck & E. S. Kappenman (Eds.), *Oxford handbook of ERP components*. New York, NY: Oxford University Press.

Kappenman, E. S., & Luck, S. J. (2012b). Manipulation of orthogonal neural systems together in electrophysiological recordings: The MONSTER Approach to simultaneous assessment of multiple neurocognitive processes. *Schizophrenia Bulletin*, *38*, 92–102.

Keil, A. (2013). Electro- and magnetoencephalography in the study of emotion. In J. Armony & P. Vuilleumier (Eds.), *The Cambridge handbook of human affective neuroscience* (pp. 107–132). Cambridge, MA: Cambridge University Press.

Keil, A., Sabatinelli, D., Ding, M., Lang, P. J., Ihssen, N., & Heim, S. (2009) Re-entrant projections modulate visual cortex in affective perception: Directional evidence from Granger causality analysis. *Human Brain Mapping*, *30*, 532–540.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 593–602. doi: 10.1001/archpsyc.62.6.593

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*, 617–627. doi: 10.1001/archpsyc.62.6.617

Kiebel, S. J., Tallon-Baudry, C., & Friston, K. J. (2005). Parametric analysis of oscillatory activity as measured with EEG/MEG. *Human Brain Mapping*, *26*, 170–177. doi: 10.1002/hbm.20153

Koessler, L., Benhadid, A., Maillard, L., Vignal, J. P., Felblinger, J., & Vespignani, H. (2008). Automatic localization and labeling of EEG sensors (ALLES) in MRI volume. *Neuroimage*, *41*, 914–923. doi: 10.1016/j.neuroimage.2008.02.039

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*, 535–540. doi: 10.1038/nn.2303

Kristjansson, S. D., Kircher, J. C., & Webb, A. K. (2007). Multilevel models for repeated measures research designs in psychophysiology: An introduction to growth curve modeling. *Psychophysiology*, *44*, 728–736. doi: 10.1111/J.1469-8986.2007.00544.X

Lantz, G., Grave de Peralta, R., Spinelli, L., Seeck, M., & Michel, C. M. (2003). Epileptic source localization with high density EEG: How many electrodes are needed? *Clinical Neurophysiology*, *114*, 63–69. doi: 10.1016/S1388-2457(02)00337-1

Le, J., Lu, M., Pellouchoud, E., & Gevins, A. (1998). A rapid method for determining standard 10/10 electrode positions for high resolution EEG studies. *Electroencephalography and Clinical Neurophysiology*, *106*, 554–558. doi: 10.1016/S0013-4694(98)00004-2

Lemieux, L., Allen, P. J., Franconi, F., Symms, M. R., & Fish, D. R. (1997). Recording of EEG during fMRI experiments: Patient safety. *Magnetic Resonance Medicine*, *38*, 943–952.

Lieberman, M. D., Berkman, E. T., & Wager, T. D. (2009). Correlations in social neuroscience aren't voodoo: Commentary on Vul et al. (2009). *Perspectives on Psychological Science*, *4*, 299–307.

Lin, C. T., Liao, L. D., Liu, Y. H., Wang, I. J., Lin, B. S., & Chang, J. Y. (2011). Novel dry polymer foam electrodes for long-term EEG measurement. *IEEE Transactions in Biomedical Engineering*, *58*, 1200–1207. doi: 10.1109/TBME.2010.2102353

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Acadademy of Sciences of the U S A*, *94*, 10979–10984.

Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology*, *49*, 549–565. doi: 10.1111/j.1469-8986.2011.01320.x

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*, 203–208.

Meehl, P. E. (1971). High school yearbooks: A reply to Schwarz. *Journal of Abnormal Psychology*, *77*, 143–148.

Mensen, A., & Khatami, R. (2013). Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics. *Neuroimage*, *67*, 111–118. doi: 10.1016/j.neuroimage.2012.10.027

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, *110*, 40–48.

Miller, G. A., Elbert, T., Sutton, B. P., & Heller, W. (2007). Innovative clinical assessment technologies: Challenges and opportunities in neuroimaging. *Psychological Assessment*, *19*, 58–73. doi: 10.1037/1040-3590.19.1.58

Miller, G.A., Gratton, G., & Yee, C.M. (1988). Generalized implementation of an eye movement correction algorithm. *Psychophysiology*, *25*, 241–243.

Miller, G. A., Lutzenberger, W., & Elbert, T. (1991). The linked-reference issue in EEG and ERP recording. *Journal of Psychophysiology*, *5*, 273–276.

Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, *35*, 99–115.

Mosher, J. C., Leahy, R. M., & Lewis, P. S. (1999). EEG and MEG: Forward solutions for inverse methods. *IEEE Transactions in Biomedical Engineering*, *46*, 245–259.

Mullinger, K., & Bowtell, R. (2011). Combining EEG and fMRI. *Methods in Molecular Biology*, *711*, 303–326. doi: 10.1007/978-1-61737-992-5_15

Ng, W. C., Seet, H. L., Lee, K. S., Ning, N., Tai, W. X., & Sutedja, M. (2009). Micro-spike EEG electrode and the vacuum-casting technology for mass production. *Journal of Materials Processing Technology*, *209*, 4434–4438.

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., & Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology*, *115*, 2292–2307. doi: 10.1016/j.clinph.2004.04.029

Nolte, G., & Dassios, G. (2005). Analytic expansion of the EEG lead field for realistic volume conductors. *Physics in Medicine and Biology*, *50*, 3807–3823.

Nunez, P. L., & Silberstein, R. B. (2000). On the relationship of synaptic activity to macroscopic measurements: Does co-registration of EEG with fMRI make sense? *Brain Topography*, *13*, 79–96.

Nuwer, M. R., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guerit, J. M., & Hinrichs, H. (1999). IFCN standards for digital recording of clinical EEG. The International Federation of Clinical Neurophysiology. [Supplement]. *Electroencephalography and Clinical Neurophysiology*, *52*, 11–14.

Onton, J., Westerfield, M., Townsend, J., & Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neuroscience and Biobehavioral Reviews*, *30*, 808–822. doi: 10.1016/j.neubiorev.2006.06.007

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*, 713–719. doi: 10.1016/S1388-2457(00)00527-7

Perrin, F., Pernier, J., Bertrand, O., Giard, M. H., & Echallier, J. F. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalography and Clinical Neurophysiology*, *66*, 75–81.

Picton, T. W., Alain, C., Woods, D. L., John, M. S., Scherg, M., & Valdes-Sosa, P. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiology and Neurootology*, *4*, 64–79.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., & Johnson, R., Jr. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*, 127–152.

Pivik, R. T., Broughton, R. J., Coppola, R., Davidson, R. J., Fox, N., & Nuwer, M. R. (1993). Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts. *Psychophysiology*, *30*, 547–558.

Resnick, S. M. (1992). Matching for education in studies of schizophrenia. *Archives of General Psychiatry*, *49*, 246.

Roach, B. J., & Mathalon, D. H. (2008). Event-related EEG time-frequency analysis: An overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophrenia Bulletin*, *34*, 907–926. doi: 10.1093/schbul/sbn093

Russell, G. S., Jeffrey Eriksen, J. K., Poolman, P., Luu, P., & Tucker, D. M. (2005). Geodesic photogrammetry for localizing sensor positions in dense-array EEG. *Clinical Neurophysiology*, *116*, 1130–1140. doi: 10.1016/j.clinph.2004.12.022

SAMHSA. (2012). *Results from the 2011 National Survey on Drug Use and Health: Summary of national findings* (NSDUH Series H-44, HHS Publication No. (SMA) 12-4713). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Scherg, M., & Von Cramon, D. (1986). Evoked dipole source potentials of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology*, *65*, 344–360.

Spencer, K. M., Dien, J., & Donchin, E. (1999). A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology*, *36*, 409–414.

Srinivasan, R., Nunez, P. L., & Silberstein, R. B. (1998). Spatial filtering and neocortical dynamics: Estimates of EEG coherence. *IEEE Transactions in Biomedical Engineering*, *45*, 814–826. doi: 10.1109/10.686789

Stam, C. J., Nolte, G., & Daffertshofer, A. (2007). Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Human Brain Mapping*, *28*, 1178–1193. doi: 10.1002/hbm.20346

Steinstrater, O., Sillekens, S., Junghoefer, M., Burger, M., & Wolters, C. H. (2010). Sensitivity of beamformer source analysis to deficiencies in forward modeling. *Human Brain Mapping*, *31*, 1907–1927. doi: 10.1002/hbm.20986

Taheri, B. A., Knight, R. T., & Smith, R. L. (1994). A dry electrode for EEG recording. *Electroencephalography and Clinical Neurophysiology*, *90*, 376–383.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, *3*, 151–162.

Tenke, C. E., & Kayser, J. (2005). Reference-free quantification of EEG spectra: Combining current source density (CSD) and frequency principal components analysis (fPCA). *Clinical Neurophysiology*, *116*, 2826–2846. doi: 10.1016/j.clinph.2005.08.007

Tierney, A. L., Gabard-Durnam, L., Vogel-Farley, V., Tager-Flusberg, H., & Nelson, C. A. (2012). Developmental trajectories of resting EEG power: An endophenotype of autism spectrum disorder. *PLoS One*, *7*, e39127. doi: 10.1371/journal.pone.0039127 PONE-D-12-01143

Tucker, D. M. (1993). Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*, *87*, 154–163.

Urbach, T. P., & Kutas, M. (2002). The intractability of scaling scalp distributions to infer neuroelectric sources. *Psychophysiology*, *39*, 791–808. doi: 10.1017/S0048577202010648

Urbach, T. P., & Kutas, M. (2006). Interpreting event-related brain potential (ERP) distributions: Implications of baseline potentials and variability with application to amplitude normalization by vector scaling. *Biological Psychology*, *72*, 333–343. doi: 10.1016/j.biopsycho.2005.11.012

Vasey, M. W., & Thayer, J. F. (1987). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: A multivariate solution. *Psychophysiology*, *24*, 479–486.

Viola, F. C., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., & Debener, S. (2009). Semi-automatic identification of independent components representing EEG artifact. *Clinical Neurophysiology*, *120*, 868–877. doi: 10.1016/j.clinph.2009.01.015

Viviani, R. (2010). Unbiased ROI selection in neuroimaging studies of individual differences. *Neuroimage*, *50*, 184–189. doi: 10.1016/j.neuroimage.2009.10.085

von Helmholtz, H. (1853). Ueber einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern mit Anwendung auf die thierisch elektrischen Versuche [On the laws of electrical current distribution in volume conductors, with an application to animal-electric experiments]. *Annalen der Physik*, *165*, 211–233.

Voytek, B., D'Esposito, M., Crone, N., & Knight, R. T. (2013). A method for event-related phase/amplitude coupling. *Neuroimage*, *64*, 416–424. doi: 10.1016/j.neuroimage.2012.09.023

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, *4*, 274–290.

Wasserman, S., & Bockenholt, U. (1989). Bootstrapping: Applications to psychophysiology. *Psychophysiology*, *26*, 208–221.

Wechsler, H., & Nelson, T. F. (2008). What we have learned from the Harvard School of Public Health College Alcohol Study: Focusing attention on college student alcohol consumption and the environmental conditions that promote it. *Journal of Studies on Alcohol and Drugs*, *69*, 481–490.

Yuval-Greenberg, S., Tomer, O., Keren, A. S., Nelken, I., & Deouell, L. Y. (2008). Transient induced gamma-band response in EEG as a manifestation of miniature saccades. *Neuron*, *58*, 429–441.

Zayas, V., Greenwald, A. G., & Osterhout, L. (2010). Unintentional covert motor activations predict behavioral effects: Multilevel modeling of trial-level electrophysiological motor activations. *Psychophysiology*, *48*, 208–217. doi: 10.1111/j.1469-8986.2010.01055.x

Zinbarg, R. E., Suzuki, S., Uliaszek, A. A., & Lewis, A. R. (2010). Biased parameter estimates and inflated Type I error rates in analysis of covariance (and analysis of partial variance) arising from unreliability: Alternatives and remedial strategies. *Journal of Abnormal Psychology*, *119*, 307–319. doi: 10.1037/a0017552

Appendix

Authors' Checklist

The checklist is intended to facilitate a brief overview of important guidelines, sorted by topic. Authors may wish to use it prior to submission, to ensure that the manuscript provides key information.

Hypotheses	YES	NO
Specific hypotheses and predictions for the electromagnetic measures are described in the introduction		

Participants	YES	NO
Characteristics of the participants are described, including age, gender, education level, and other relevant characteristics		

Recording characteristics and instruments	YES	NO
The type of EEG/MEG sensor is described, including make and model		
All sensor locations are specified, including reference electrode(s) for EEG		
Sampling rate is indicated		
Online filters are described, specifying the type of filter and including roll-off and cut-off parameters (in dB, or by indicating whether the cut-off represents half-power/half amplitude)		
Amplifier characteristics are described		
Electrode impedance or similar information is provided		

Stimulus and timing parameters	YES	NO
Timing of all stimuli, responses, intertrial intervals, etc., are fully specified; ensure clarity that intervals are from onset or offset		
Characteristics of the stimuli are described such that replication is possible		

Description of data preprocessing steps	YES	NO
The order of all data preprocessing steps is included		
Rereferencing procedures (if any) are specified, including the location of all sensors contributing to the new reference		
Method of interpolation (if any) is described		
Segmentation procedures are described, including epoch length and baseline removal time period		
Artifact rejection procedures are described, including the type and proportion of artifacts rejected		
Artifact correction procedures are described, including the procedure used to identify artifacts, the number of components removed, and whether they were performed on all subjects		
Offline filters are described, specifying the type of filter and including roll-off and cut-off parameters (in dB, or by indicating whether the cut-off represents half-power/half amplitude)		
The number of trials used for averaging (if any) is described, reporting the number of trials in each condition and each group of subjects. This should include both the mean number of trials in each cell and the range of trials included		

Measurement procedures	YES	NO
Measurement procedures are described, including the measurement technique (e.g., mean amplitude), the time window and baseline period, sensor sites, etc.		
For peak amplitude measures, the following is included: whether the peak was an absolute or local peak, whether visual inspection or automatic detection was used, and the number of trials contributing to the averages used for measurement		
An a priori rationale is given for the selection of time windows, electrode sites, etc.		
Both descriptive and inferential statistics are included		

Statistical analyses	YES	NO
Appropriate correction for any violation of model assumptions is implemented and described (e.g., Greenhouse-Geisser, Huynh-Feldt, or similar adjustment)		
The statistical model and procedures are described and results are reported with test statistics, in addition to p values		
An appropriate adjustment is performed for multiple comparisons		
If permutation or similar techniques are applied, the number of permutations is indicated together with the method used to identify a threshold for significance		

Figures

	YES	NO
Data figures for all relevant comparisons are included		
Line plots (e.g., ERP/ERF waveforms) include the following: sensor location, baseline period, axes at zero points (zero physical units and zero ms), appropriate *x*- and *y*-axis tick marks at sufficiently dense intervals, polarity, and reference information, where appropriate		
Scalp topographies and source plots include the following: captions and labels including a key showing physical units, perspective of the figure (e.g., front view; left/right orientation), type of interpolation used, location of electrodes/sensors, and reference		
Coherence/connectivity and time-frequency plots include the following: a key showing physical units, clearly labeled axes, a baseline period, the locations from which the data were derived, a frequency range of sufficient breadth to demonstrate frequency-specificity of effects shown		

Spectral analyses

	YES	NO
The temporal length and type of data segments (single trials or averages) entering frequency analysis are defined		
The decomposition method is described, and an algorithm or reference given		
The frequency resolution (and the time resolution in time-frequency analyses) is specified		
The use of any windowing function is described, and its parameters (e.g., window length and type) are given		
The method for baseline adjustment or normalization is specified, including the temporal segments used for baseline estimation, and the resulting unit		

Source-estimation procedures

	YES	NO
The volume conductor model and the source model are fully described, including the number of tissues, the conductivity values of each tissue, the (starting) locations of sources, and how the sensor positions are registered to the head geometry		
The source estimation algorithm is described, including all user-defined parameters (e.g., starting conditions, regularization)		

Principle component analysis (PCA)

	YES	NO
The structure of the EEG/MEG data submitted to PCA is fully described		
The type of association matrix is specified		
The PCA algorithm is described		
Any rotation applied to the data is described		
The decision rule for retaining/discarding PCA components is described		

Independent component analysis (ICA)

	YES	NO
The structure of the EEG/MEG data submitted to ICA is described		
The ICA algorithm is described		
Preprocessing procedures, including filtering, detrending, artifact rejection, etc., are described		
The information used for component interpretation and clustering is described		
The number of components removed (or retained) per subject is described		

Multimodal imaging

	YES	NO
Single-modality results are reported		

Current source density and Laplacian transformations

	YES	NO
The algorithm used and the interpolation functions are described		

Single-trial analyses

	YES	NO
All preprocessing steps are described		
A mathematical description of the algorithm is included or a reference to a complete description is provided		