Preregistration: A Solution to Undisclosed Analytic Flexibility in ERP Research

Mariella Paul[a,b,c,§], Gisela H. Govaart[a,b,d,§], Antonio Schettino[e,f,§,*]

[a] Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraβe 1a, 04103 Leipzig, Germany

[b] Humboldt-Universität zu Berlin, Faculty of Philosophy, Berlin School of Mind and Brain, Luisenstraße 56, 10117 Berlin, Germany

[c] Psychology of Language Department, University of Göttingen, Goßlerstraße 14, 37073 Göttingen, Germany

[d] Charité – Universitätsmedizin Berlin, Einstein Center for Neurosciences Berlin, 10117, Berlin, Germany

[e] Erasmus Research Services, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands

[f] Institute for Globally Distributed Open Research and Education (IGDORE), Ubud (Indonesia)

[§] Equal contribution

[*] Corresponding author: Antonio Schettino, Erasmus Research Services, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, email: schettino@eur.nl

Highlights

- Undisclosed analytic flexibility increases false positives and inflates effect sizes
- Relevant for EEG research: many preprocessing/analysis pipelines, low standardization
- To limit analytic flexibility: *Preregister* research plan before data collection
- How to preregister data preprocessing and analysis steps in typical ERP studies
- Benefits for individuals and scientific discipline outweigh costs

Abstract

Confirmation bias and pressure to publish may prompt the (unconscious) exploration of various methodological options and reporting only the ones that lead to a (statistically) significant outcome. This *undisclosed analytic flexibility* increases false positive results and inflates effect sizes, ultimately creating a skewed representation of knowledge. This issue is particularly relevant in EEG research, where a myriad of preprocessing and analysis pipelines can be used to extract information from complex multidimensional data. One solution to limit undisclosed analytic flexibility is *preregistration*: researchers write a time-stamped, publicly accessible research plan with hypotheses, data collection plan, and intended preprocessing and statistical analyses before the start of a research project. In this manuscript, we present an overview of the problems associated with undisclosed analytic flexibility (particularly in human neuroimaging and electrophysiology), discuss why and how EEG researchers would benefit from adopting preregistration, provide guidelines and examples on how to preregister data preprocessing and analysis steps in typical ERP studies, and conclude by addressing common rebuttals as well as clarifying possibilities and limitations of this open science practice.

KEYWORDS: EEG, ERP, open science, preregistration

1. Introduction

In his essay entitled *Reflections on the decline of science in England, and on some of its causes* (1830), Charles Babbage identifies several questionable behaviors that scientists may engage in, for example *forging* (“*the forger is one who, wishing to acquire a reputation for science, records observations which he has never made*”; p. 177) and *trimming* (“*clipping off little bits here and there from those observations which differs most in excess from the mean, and in sticking them on to those which are too small”*; p. 178). *Cooking* (“*an art of various forms, the object of which is to give to ordinary observations the appearance and character of those of the highest degree of accuracy*”; p. 178) refers to selectively reporting only observations that support a preferred conclusion.[1] An example of cooking is *cherry-picking* (“*to make multitudes of observations, and out of these to select those only which agree, or very nearly agree”*; p. 178). This fallacy is an effective weapon in the hands of *confirmation bias*, the tendency to search, interpret, and remember information that supports prior beliefs while ignoring evidence against them (Nickerson, 1998). Left unchecked, confirmation bias limits our understanding of complex psychological, social, and economic phenomena (Board of Governors of the Federal Reserve System, 2015; Camerer et al., 2016; Open Science Collaboration, 2015), undermines the validity and integrity of biomedical research (Hannink et al., 2013; Jones et al., 2017; Trinquart et al., 2018), leads to

[1] Selective reporting needs not be committed intentionally or with malicious intent. Researchers may just be following procedures currently accepted in their field, sometimes without critically thinking about their appropriateness (e.g., Gigerenzer, 2004). A prominent example is a study reporting alleged evidence of precognition abilities in humans (Bem, 2011), which served as a stark reminder of the widespread adoption of suboptimal research practices among (social) scientists (John et al., 2012; Wagenmakers et al., 2011). This (and other events; e.g., Levelt et al., 2012) motivated mainstream discussions on how to improve study transparency (Nosek and Bar-Anan, 2012; Simmons et al., 2012; Simonsohn, 2013; Steegen et al., 2016; Wilson et al., 2017), theoretical and methodological rigor (Devezer et al., 2020; Oberauer and Lewandowsky, 2019), statistical literacy (Cumming, 2014; Kruschke and Liddell, 2017; Wasserstein and Lazar, 2016), and incentive structures (Nosek et al., 2012).

considerable waste of resources (Chalmers et al., 2014; Vedula et al., 2012), and may threaten people's lives (Anand et al., 2014; Topol, 2004).

1.1 Undisclosed Analytic Flexibility

Besides selectively picking observations that conform to a preferred hypothesis, *undisclosed analytic flexibility* is another way to "cook" results. Data collected in an experimental or observational study are rarely interpreted in their raw form. Instead, researchers typically apply a series of transformations to deal with outliers and missing data (Enders, 2010; Hawkins, 1980), combine or discretize variables into composite indices, change the unit of measurement, and so on. In other words, "*data are to a certain extent actively constructed*" (Steegen et al., 2016, p. 702). Moreover, there are countless statistical techniques that can be chosen to analyze the preprocessed data, including classical null hypothesis tests (Field et al., 2012; Judd et al., 2017a) and their robust counterparts (Wilcox, 2016), Bayesian parameter estimation (Kruschke, 2014; McElreath, 2018), and more. This myriad of choices that a researcher has to make during the research process is referred to as *analytic flexibility*.[2] Importantly, selecting data preprocessing and analysis pipelines is, in many cases, neither univocal nor objective, and its rationale is typically lacking: it is rare to encounter papers with a clear justification of the threshold used to define outliers (e.g., why use a cutoff of 2.5 standard deviations instead of 3?), or the employed statistical technique (e.g., why a classical ANOVA instead of its robust bootstrapped counterpart?). This should

[2] Analytic flexibility is sometimes also called *researcher degrees of freedom* (Simmons et al., 2011) or *garden of forking paths* (Gelman and Loken, 2013). However, these terms may refer to *all* choices that researchers make throughout their workflow, including hardware and software selected for data collection and analysis, the type and number of stimuli presented to participants, and much more. These choices can have tangible consequences on study results and interpretation. For instance, the same analysis pipeline on the same dataset can lead to quantifiably different results when run with different software (Bowring et al., 2019; see also Eklund et al., 2016). Typically, larger and more homogeneous samples of both participants and stimuli increase statistical power (Judd et al., 2017b), and ignoring these sources of variability in the applied statistical model has a direct impact on the generalizability of the results from a particular dataset to other (hypothetically similar) scenarios (Yarkoni, 2019). Throughout this paper, we limit our discussion to analytic flexibility during the preprocessing and analysis phases of the research cycle.

not necessarily be ascribed to blind following of statistical rituals (Gigerenzer, 2004), but may be due to the fact that there are multiple reasonable processing steps that can be applied to the same dataset (Steegen et al., 2016).

Undisclosed analytic flexibility occurs when methodological choices on preprocessing pipelines and statistical analysis made *after* seeing the data go unreported. This has a profound impact on the results, since it has been shown to increase false positives, i.e., the incorrect rejection of a null hypothesis, from the nominal 5% (assuming $\alpha = 0.05$) to over 50% (Ioannidis, 2005; Simmons et al., 2011). Also, estimated effect sizes are remarkably inflated, a phenomenon known as “the winner’s curse” (Ioannidis, 2008; Lane and Dunlap, 1978): large-scale replication projects in a sample of published studies in psychology and economics showed that the mean effect size of the replications was around half that of the original effects (Camerer et al., 2016; Open Science Collaboration, 2015), and an overestimation of similar magnitude has been calculated in genetic association studies (Xiao and Boehnke, 2009). The repercussions are troublesome: undisclosed analytic flexibility limits our ability to precisely detect genetic variants that predispose to human disease (Grinde et al., 2017; Palmer and Pe’er, 2017), develop effective evidence-based clinical treatment (Lamberink et al., 2018) and, more generally, build cumulative knowledge in favor or against competing theories when following an idealized hypothetico-deductive model of knowledge accrual.

To summarize, undisclosed analytic flexibility can lead to: (i) claiming that a particular effect is present in the data when, in fact, it may not be; and (ii) overestimating its size and, consequently, its practical usefulness. This issue is not only present in psychology, economics, epidemiology, and genetics, but also in human neuroimaging and electrophysiology.

1.2 Undisclosed Analytic Flexibility in Human Neuroimaging Research

Analytic flexibility in neuroimaging – particularly functional magnetic resonance imaging (fMRI) – research has received considerable attention over the last years. A widely discussed study by Vul et al. (2009) highlighted implausibly high brain-behavior correlations in social and affective neuroscience, ascribed to *double-dipping* – selecting voxels based on a functional analysis and then reporting the results of the same analysis from just the selected voxels (Kriegeskorte et al., 2009) – as well as low statistical power and effect size inflation (Yarkoni, 2009). A review of papers published in a number of cognitive neuroscience journals between 2011 and 2014 (Szucs and Ioannidis, 2017) reported very high effect sizes (d = 0.34 – 1.22), likely inflated considering the overall low sample size (df = 10 – 28) and statistical power (median of 11%, 40%, and 70%, assuming benchmark effect sizes of d = 0.2, d = 0.5, and d = 0.8, respectively). These results are congruent with the findings of a previous meta-analysis reporting a median statistical power of 8% in a sample of human neuroimaging studies published between 2006 and 2009 (Button et al., 2013). Importantly, undisclosed analytic flexibility has been identified as a crucial factor in determining overoptimistic effect size estimates and increased false positives in this literature (Button et al., 2013, p. 367; Ioannidis, 2005, p. 0698; Szucs and Ioannidis, 2017, p. 13).

Recent studies have directly examined the impact of analytic flexibility on the results reported in the fMRI literature. Carp (2012a) surveyed how methods were reported in a sample of 241 neuroimaging studies published between 2007 and 2011 and identified 207 unique combinations of analysis pipelines, which could rarely be fully reproduced due to frequent omissions of important methodological details. A follow-up study (Carp, 2012b)

assessed the influence of 6,912 unique preprocessing and analysis pipelines on the variability of statistical maps obtained from a single dataset, revealing broad variability in activation strength as well as localization of peak activation.[3] Similarly, in a recent multi-laboratory study (Botvinik-Nezer et al., 2020), 70 independent research teams conducted a whole-brain analysis on the same fMRI dataset using their preferred procedures, and reported whether each of 9 prespecified hypotheses were statistically supported. Each team employed different preprocessing and analysis pipelines. Across all hypotheses, 20% of teams reported a result that differed from the majority of all other teams (note that the largest possible variation is 50%, indicating maximally inconsistent results across teams). Prediction markets further revealed that researchers in the field were not able to successfully predict the results: instead, they would systematically overestimate the probability of each hypothesis being statistically confirmed, whether or not they participated in the analysis themselves.

These studies highlight the large number of analytic options available to researchers. If not transparently reported, this kind of flexibility has a major effect on the results reported in the neuroimaging literature which, in turn, inform the validity of specific hypotheses and theoretical frameworks. Of note, most analysis choices used in the studies mentioned above are justifiable. While some researchers might engage in "method shopping" – selecting procedures that provide a higher probability of statistically significant findings at a potential cost of increased error rates (Poldrack et al., 2017, p. 8) –, here we do not question whether such choices are appropriate *per se*, but rather highlight the importance of transparently documenting them as part of the research workflow.

[3] Similar findings have been reported in positron emission tomography (PET) research. A recent study analyzing the impact of 384 commonly used preprocessing pipelines on a PET dataset showed that only 36% of these preprocessing strategies replicated the statistically significant results reported in the original paper (Nørgaard et al., 2019).

1.3 Undisclosed Analytic Flexibility in Human Electrophysiology Research

Researchers analyzing magneto- and electroencephalographic data (M/EEG) face similar challenges to what has been described above for fMRI. M/EEG signals are complex and multidimensional: space, time, and frequency – assessed via indices such as activity magnitude, connectivity, and network properties (Kida et al., 2016) – interact with experimental designs of various complexity, often resulting in a large number of independent and dependent variables.

In addition, the raw signal recorded by electrodes (and magnetometers) must undergo a series of preprocessing steps that magnify cerebral activity against environmental noise (Cohen, 2014; Hansen et al., 2010; Luck, 2014). Offline modifications of the continuous EEG signal include: (*i*) re-referencing to the activity of specific electrodes or the average activity of all electrodes on the scalp; (*ii*) interpolation of noisy channels; (*iii*) high-, low-, or band-pass filtering; (*iv*) correcting or rejecting physiological artifacts (e.g., blinks, muscular activity); and (*v*) removal of baseline activity and segmentation into epochs around the event(s) of interest (Luck, 2014). Needless to say, there is considerable flexibility at each of these steps: (*i*) popular reference methods include vertex, linked mastoids or ears, average reference, and Reference Electrode Standardization Technique (for reviews, see Dong et al., 2019; Liu et al., 2015), and their choice is not always obvious with respect to the experimental design or dependent variables of interest; (*ii*) channel interpolation – e.g., nearest neighbor (Shepard, 1968), thin-plate spline (Harder and Desmarais, 1972), spherical spline (Perrin et al., 1989), 3-D spline (Law et al., 1993) – is also a potential source of stochastic error (Fletcher et al., 1996), and its choice is often left to the software used for

preprocessing; (*iii*) there are many different filter types available, and there is considerable flexibility in setting the exact parameters for the filter, e.g., the cut-off frequency, transition width, etc.; moreover, filtering can severely distort the signal (e.g., Kappenman and Luck, 2010), which even led some to propose its exclusion from preprocessing pipelines in specific experimental designs (VanRullen, 2011; but see Widmann and Schröger, 2012); (*iv*) there is a large number of artifact detection, correction, and rejection techniques (for a review, see Jiang et al., 2019), each with its own expected user input (e.g., from tweaking a few parameters in a fully automated algorithm to visual inspection of epochs for manual removal); (*v*) for baseline correction, the selected time window can vary in length and location (i.e., more proximal or distal from the event of interest); also, baseline correction can bias scalp topographies (Urbach and Kutas, 2006), which may lead researchers to favor other techniques (e.g., Alday, 2019). Moreover, the *order* in which some of the above mentioned steps are performed may distort the resulting waveforms, e.g., filtering epoched instead of continuous EEG data may create edge artifacts, particularly when using inappropriate filter types or cut-off values (Luck, 2014, pp. 247–248; see also Widmann et al., 2015).

Thus, EEG researchers have a considerably long list of data transformation steps at their disposal. Recent papers directly demonstrated that analytic flexibility may influence the results and interpretation of electrophysiological data. Robbins et al. (2020) applied four preprocessing pipelines (Bigdely-Shamlo et al., 2020a; Chang et al., 2020; Winkler et al., 2011) to a large and heterogeneous EEG dataset containing 7.8 million event-related epochs (Bigdely-Shamlo et al., 2020a). There were differences in the spectral characteristics of the processed signals, attributable to the different artifact correction procedures across preprocessing pipelines. In addition, small parameter deviations in otherwise very similar artifact correction algorithms were shown to distort the signal, especially in low frequency

bands. Calculation of event-related epochs was also affected by specific steps in the selected preprocessing pipeline: for example, outlier detection algorithms may be incorporated in some pipelines (e.g., Bigdely-Shamlo et al., 2020b) but not in others.

Another example pertains to the error-related negativity (ERN). This ERP component of negative polarity peaks ~80-150 milliseconds after an erroneous motor response in speeded tasks, is largest at midline frontal and central electrode sites, and originates from the anterior cingulate cortex (Falkenstein et al., 1991; Gehring et al., 1993). A recent paper (Sandre et al., 2020) highlighted cross-study variability in the selection of reference location, baseline correction, and electrode site from which signal amplitudes were measured. The authors systematically compared 72 commonly used preprocessing pipelines to examine their effects on the resulting ERN amplitude. Results showed that different preprocessing choices had a remarkable influence on the within- and between-subject effects typically assessed in ERN research – i.e., post-error slowing and gender differences –, with mastoid reference, distal baseline correction periods (i.e., further away from the time-locked response), single electrode site, and peak-to-peak amplitude measures leading to larger estimated differences between conditions (see also Klawohn et al., 2020).

Analytic flexibility does not only occur at the level of preprocessing. Rich data sets such as M/EEG can be analyzed in multiple different ways – e.g., event-related potentials (ERPs; Luck, 2014), EEG microstates (Michel and Koenig, 2018), time-frequency (Cohen, 2014), functional connectivity (Bastos and Schoffelen, 2016), steady-state evoked potentials (Regan, 1977), source localization (Michel and He, 2019) –, with different objectives, dependent variables, and levels of analytic sophistication. Last but not least[4], researchers have

[4] Other factors, including context-dependent psychometric properties of brain measures (e.g., Clayson and Miller, 2017) and suboptimal reporting of methodological details in published papers (e.g., Clayson et al., 2019; Larson and Carbine, 2017), further complicate the picture and warrant caution on the reliability of the psychophysiological literature (Baldwin, 2017).

multiple valid options for statistical tests, such as ANOVAs and *t*-tests (Luck, 2014, chap. 10), cluster-based permutation tests (Maris and Oostenveld, 2007), Bayes factors (Keysers et al., 2020), linear mixed effects models (Frömer et al., 2018), threshold-free cluster enhancement (Mensen and Khatami, 2013), and more.

As discussed above, undisclosed analytic flexibility leads to an increase of false positives and an overestimation of effect sizes. Quoting Sandre et al. (2020, p. 35): "*[...] different ways of processing the same data can lead researchers to different conclusions, demonstrating yet again that transparency of all processing decisions is a necessity.*". We concur: transparently reporting all analytic choices would increase study reproducibility and, more generally, the trustworthiness of the electrophysiological literature. However, as mentioned before, researchers (just like other human beings) tend to shape their analytic choices with the (largely implicit) aim to confirm their prior beliefs (Nickerson, 1998), and post-hoc justification of said choices is rationalized under the "illusion of objectivity" (Kunda, 1990; Pyszczynski and Greenberg, 1987).

Such biases cannot take place if analytic choices are determined *before* the data are collected. Sandre et al. suggest that "*a single processing stream should be finalized before any analyses are undertaken*" (2020, p. 35) or, in other words, researchers should commit to a specific analytic path before exploring their data. This practice is called preregistration.

2. Preregistration

Preregistrations are time-stamped, publicly accessible documents with hypotheses, data collection plan, and intended preprocessing and statistical analyses, written *before* the

start of a research project.[5] In other words, researchers commit to one among many ways in which the study can be conducted and analyzed. This document is uploaded on a trusted online repository – e.g., Open Science Framework (*OSF*; https://osf.io/), ClinicalTrials.gov (https://clinicaltrials.gov/), American Economic Association's registry for randomized controlled trial (*AEA RCT Registry*; https://www.socialscienceregistry.org/) –, which assigns it a date and time. The protocol is made public immediately or after an embargo period. Date and time of submission ensure that the research plan was devised before starting the study.

The popularity of preregistration has skyrocketed in recent years: for example, the number of public documents uploaded on the OSF went from 38 in 2012 to 36,675 in 2019 (Bakker et al., 2018). Many journals now explicitly encourage this practice by awarding "preregistration badges" (https://osf.io/tvyxz/; see Kidwell et al., 2016), including *Psychological Science* and *Cortex* (for a full list, see https://tinyurl.com/COS-badges).

2.1 Advantages of Preregistration

The advantages of preregistering research plans are manyfold. First, preregistration allows a clear separation between confirmatory (hypothesis-driven) and exploratory (data-driven) analysis (Nosek et al., 2018; Wagenmakers et al., 2012). Using the same data to simultaneously formulate and confirm a hypothesis increases the probability of false positives and inflated effect sizes, which impacts replicability (e.g., Kriegeskorte et al., 2009).[6] Preregistration allows researchers to specify the rationale and hypotheses of the study while

[5] Preregistration at a later point in time is also possible, as long as authors transparently report at which stage of the study they crafted the protocol and declare that they are not yet aware of any results. Another possibility is to preregister analysis plans of data that have already been collected but not accessed, i.e. secondary data analysis (Mertens and Krypotos, 2019; Van den Akker et al., 2019).

[6] A clear separation between exploratory and confirmatory analyses has direct implications on the interpretability of the reported results. In a review published in the *International Journal of Epidemiology* (Swaen et al., 2001), the authors selected a subsample of articles investigating substances or work processes generally recognized as being carcinogenic to humans. Of all the factors included in the scoring procedure, the one more strongly associated with true vs. false positive outcome was whether the study had specific *a priori* hypotheses as opposed to being a "fishing expedition [*sic*]", the latter being three times more likely to yield false positive findings.

also maintaining flexibility with respect to additional exploratory analyses, provided that they are included in a different section of the final manuscript. This strategy also precludes presenting any hypotheses generated after observing the data as if they were *a priori*, or "hypothesizing after the results are known" (HARKing; Kerr, 1998). This practice is particularly difficult to identify in published papers because readers can only access what the authors reported after collecting, analyzing, and interpreting the data, without knowing whether the hypotheses described in the introduction were originally unanticipated (or even considered implausible) until reassessed in light of the collected empirical evidence.[7] In addition, preregistration can help readers identify selective outcome reporting (John et al., 2012; Simmons et al., 2011): for instance, discrepancies between planned and actual sample size, unforeseen moderators, or flexible exclusion criteria may be detected more easily and require justification. Furthermore, it has been argued (Nosek et al., 2019, 2018) that preregistration can contribute to mitigating publication bias in the academic literature (Nissen et al., 2016; Rosenthal, 1979; Scargle, 2000), since research plans are discoverable regardless of whether the final report is ultimately published in peer-reviewed journals. Yet, in our opinion, publication bias can only be effectively mitigated when *results* are published regardless of study outcome, that is, via Registered Reports (see *Section 4.1*). Nonetheless, discoverability of research plans is a useful step in making the entire research process discoverable. Importantly, preregistration is not only helpful when hypotheses are tested or *p*-values are reported (McPhetres, 2020), but also for exploratory[8] and qualitative research (Dirnagl, 2020; Haven and Van Grootel, 2019) and when using other statistical procedures

[7] This problem is magnified by the fact that, at least in some areas in psychology, theoretical frameworks and research hypotheses are often underspecified, which decreases their explanatory power and predictive utility (Meehl, 1967; Szollosi and Donkin, 2019a).

[8] A preregistration can serve as a 'log' for exploratory research, to make the many choices during the research process transparent: "Methodological and analytic flexibility is maintained but disclosed." (Dirnagl, 2020, p. 4).

(e.g., specify and justify in advance what priors will be used in a planned Bayesian analysis; see Depaoli and van de Schoot, 2017). The main objective is to achieve as much transparency as possible at all stages of the research cycle.

Correlational evidence accumulated over the past 20 years in several disciplines suggests that preregistration may facilitate the publication of non-significant findings, thus providing a more accurate representation of available knowledge. For instance, Kaplan and Irvin (2015) reviewed a sample of randomized clinical trials funded by the National Heart, Lung, and Blood Institute evaluating drugs or dietary supplements for the treatment or prevention of cardiovascular disease. Of the 55 selected studies, 30 were published before and 25 after the year 2000, when study registration on ClinicalTrials.gov became compulsory in the U.S. following the Food and Drug Administration Modernization Act in 1997. Results showed that 57% of the studies published before 2000 showed a significant benefit of the intervention, as opposed to only 8% of trials published after 2000.[9] Similar results were reported in a (preregistered) meta-analysis of meta-analyses of orthodontics and dentofacial orthopedics studies: registered trials reported less favorable intervention effects compared to unregistered trials (Papageorgiou et al., 2018). Preregistration can also help identify whether funding sources are correlated with study outcome, potentially uncovering questionable practices due to (undisclosed) conflicts of interest. For instance, a review of studies of safety and efficacy trials for a wide array of drugs (Bourgeois et al., 2010) revealed that trials funded by industry were less likely to be published within 2 years from study completion and

[9] The clinical studies considered in Kaplan and Irvin (2015) were specifically chosen to be large, well-funded projects, likely to get published even if results were not statistically significant. Thus, their work does not directly show that preregistered studies are easier to get published. Yet, it does suggest that, if studies are preregistered, non-significant findings are more likely to be reported as such, instead of being *p*-hacked to chase publication.

most likely to report a positive outcome (85%, as opposed to 50% for government-funded trials).

2.2 Career Benefits Associated with Preregistration

Besides being advantageous for whole research fields, anecdotal experience and preliminary evidence suggest that preregistration can be beneficial for individual researchers as well (Allen and Mehler, 2019; McKiernan et al., 2016; Toth et al., 2020; Wagenmakers and Dutilh, 2016). Generally speaking, drafting a thorough preregistration – preferably with the help of useful templates and checklists (see *Section 3*) – can improve the experimental design not only because authors are stimulated to think more carefully about the research plan, but also because feedback from peers can be solicited early and incorporated when most valuable, that is, when there is still time to make changes. Early-career researchers (ECRs) may benefit even more from learning this skill, since they are often directly involved with the ideation and development of the research project, data preprocessing and analysis, and writing of the final report. Preregistering a study as an ECR can also give a stronger sense of ownership over ideas that were originally conceptualized by their supervisors, for example by having a clearer overview on the different steps of the workflow, making informed decisions about the rationale, experimental design, and planned analyses early on in the project.

Given an increasing interest in transparency, we expect ECRs to be working in an environment that values – and even requires – a certain level of commitment to open science practices, of which preregistration is an example. As mentioned earlier, a growing number of journals encourage preregistration, e.g., by means of badges. We surmise that the more preregistered papers will appear in the literature, the more central their role in the credibility

of confirmatory research will become. In academia[10], funding agencies appreciate the importance of study preregistration in medical and non-medical disciplines: for example, the recent COVID-19 Programme by the Dutch funder *ZonMw* includes specific open science guidelines[11] for prospective applicants, among which mandatory preregistration of animal studies and "strongly recommended" preregistration for all other studies. Last but not least, researchers who preregister their studies may be perceived as more trustworthy, because they are willing to open all products of their workflow to their peers for scrutiny. However, the only study (to our knowledge) that empirically investigated whether preregistration increases peers' trust in the final publication revealed inconclusive evidence either in favor or against this hypothesis (S. M. Field et al., 2020), leaving this question open for future examinations.

The advantages of preregistration in neuroimaging and electrophysiology have not yet systematically been evaluated. Nonetheless, the data accumulated in other disciplines provide a number of insights, practical examples, and learned lessons that can guide a widespread and informed implementation of this practice in our research fields. When done properly, preregistration works as intended.

3. Recommendations for Preregistration of ERP research

In this section, we provide guidelines on how to transparently document the planned analytic choices in a preregistration of a prototypical ERP study. We focus on ERPs because of their widespread use in cognitive and clinical research (Hajcak et al., 2019; Helfrich and

[10] Solid project management skills are also extremely valuable outside of academia, where careful planning can help prioritize goals in a fast-paced environment (see Powell, 2018).
[11] https://tinyurl.com/ZonMw-COVID19-OS

Knight, 2019), although most of these recommendations can still be useful when using other signal processing techniques (e.g., ERP and time-frequency analyses have many preprocessing steps in common). Furthermore, we only include sections that would decrease researchers' flexibility during signal preprocessing and statistical analysis. As discussed in *Section 1.3*, these steps are complex and multifaceted, with many reasonable choices that can lead to qualitatively different interpretations of the data; therefore, transparently documenting them would significantly lower the risk of false positives. However, other aspects of a study should also be carefully planned and included in the preregistration protocol, e.g., the rationale behind the chosen sample size (including a power analysis), inclusion and exclusion criteria, and stimulus characteristics.

We encourage researchers to craft a document that is *specific*, *precise*, and *exhaustive* (Veldkamp, 2017, chap. 6; Wicherts et al., 2016). A preregistration is *specific* when it includes a detailed description of all phases of the research workflow, from the initial design of the study to the information reported in the final manuscript; *precise* when the research plan is interpretable in only one way (e.g., there is no ambiguity in what statistical model will be fit to the collected data); and *exhaustive* when the research plan excludes the possibility of deviations from the preregistered protocol (see also McPhetres, 2020).

In our experience, it is very useful to run a pilot study before drafting the preregistration document. Advantages include: (*i*) gauge the feasibility of recruitment, randomization, and assessment procedures, especially when testing clinical populations and/or evaluating a novel treatment (e.g., Leon et al., 2011); (*ii*) ensure that task instructions are clear for participants; (*iii*) confirm that the target ERP component(s) are elicited; (*iv*) test preprocessing and analysis pipelines for possible bugs, errors, and/or computational

feasibility. We emphasize that small scale pilot studies should *not* be used to estimate effect sizes to inform *a priori* power analysis (Kraemer et al., 2006; Lakens and Albers, 2017).

3.1 Preregistration templates

In principle, any time-stamped, accessible protocol with a clear study plan can serve as a preregistration. However, ready-made templates can greatly facilitate the inclusion of preregistration in researchers' workflows by providing a list of bullet points (Bakker et al., 2018; Wicherts et al., 2016). In addition, hosting preregistrations on online platforms that are popular among the research community (rather than, for example, personal websites) can improve accessibility. One of the most popular platforms is the OSF, which offers several preregistration templates[12] differing on topic, length, and specificity. While less extensive templates (e.g., AsPredicted: https://aspredicted.org/) are typically used by newcomers for their first preregistration, we would rather recommend the standard OSF Prereg template, whose increased level of detail facilitates the creation of specific, precise, and exhaustive preregistrations that more effectively decrease the risk of undisclosed analytic flexibility. A template specifically for preregistration of EEG studies was started during a hackathon at the annual meeting of the Society for the Improvement of Psychological Science (SIPS) in 2019 (Algermissen et al., 2019) and is currently being developed online by an active community of volunteers. Readers are welcome to contribute to (and use) the current draft at https://tinyurl.com/eegprereg.

Below we provide some examples on how to preregister typical preprocessing and analysis steps in an ERP study. Please note that a good preregistration should also be explicit about the *order* of preprocessing steps: the examples below are for illustrative purposes and should be adjusted based on the pipeline that is appropriate for the specific study. Moreover,

[12] https://osf.io/zab38/wiki/home/

the preregistration should be specific about the software and (standardized) pipelines that will be used to carry out the preprocessing steps. If researchers plan to use the default settings of a given software, they should also include its version number, since these settings might change with different versions.

3.2 Preprocessing

3.2.1 Re-referencing

A preregistration should specify which electrodes will be chosen for the offline re-referencing. Common offline reference channels include the linked mastoids, ears, vertex, or an average reference. An example could be: *The data will be re-referenced offline to the mean of the mastoids.*

3.2.2 Filtering

Preregistering the parameter values of the filters that will be applied to the recorded EEG data should be detailed enough to theoretically allow readers to completely reproduce each filter (see Widmann et al., 2015). This includes specifying not only the filter cut-off frequency, but also the type (e.g., butterworth, finite impulse response, infinite impulse response), transition width, passband edge, the order for the transition bandwidth, at what point during the preprocessing pipeline the filter was applied (e.g., to continuous or segmented data), and – in case of multiple filters – the order in which the filters are applied. For example (modified from Schindler et al., 2018, p. 9): *The continuous EEG data will be filtered with separate Hamming windowed sinc finite impulse response (FIR) filters (Widmann, 2006): (1) high-pass: passband edge 0.5 Hz, filter order 1,690, transition*

bandwidth 0.5 Hz, cutoff frequency (−6 dB) 0.25 Hz; (2) low-pass: passband edge 30 Hz, filter order 114, transition bandwidth 7.4 Hz, cutoff frequency (−6 dB) 33.71 Hz.

3.2.3 Trial segmentation and time-locking

For trial segmentation, it is especially important to specify *when* the continuous EEG data are segmented, because this has implications for other preprocessing steps.[13] When preregistering trial segmentation, it is also important to specify what the trial will be time-locked to – e.g., stimulus onset, participant's motor response – and how long the pre- and post-stimulus period will be. For example: *We will time-lock trials to the onset of the word, with a pre-stimulus period of 200 ms and a post-stimulus period of 1000 ms.*

3.2.4 Interpolation

With interpolation, the EEG signal recorded from noisy channels is replaced with estimated activity from neighboring electrodes. In the preregistration protocol, one should prespecify the criteria used to identify noisy channels as well as the algorithm that will be used for the interpolation. For example (modified from Schindler et al., 2018, p. 9): *Channels will be interpolated via spherical spline procedure (Perrin et al., 1989), as implemented in EEGLAB v14.1.1 (Delorme and Makeig, 2004), if one or more of the following values exceeds a z-score threshold of ±3: variance, mean correlation, and Hurst exponent (see Nolan et al., 2010).*

[13] As already mentioned, high-pass filters may produce edge artifacts when applied to non-continuous signal; thus, segmentation is typically performed after these filters are applied. Other methods are also effective, e.g., extending epochs with zeroes (zero-padding) so that edge artifacts do not affect the signal of interest.

3.2.5 Artifact rejection and correction

Approaches to artifact rejection can be roughly divided into three categories: (1) automatic; (2) manual; and (3) semi-automatic, that is, a combination of automatic and manual approaches. Many different algorithms for automatic artifact rejection are available and implemented in different software and toolboxes. For example, the FASTER algorithm (Nolan et al., 2010) calculates various statistical parameters of the signal and defines data as artifactual when exceeding a pre-specified *z*-score (e.g., ±3). This approach can be additionally integrated by identifying the frequency bands in which one would expect artifacts to occur, e.g., 110 to 140 Hz for muscle artifacts (see Delorme et al., 2007). From a computational perspective, fully automatic approaches are more reproducible, although their sensitivity and specificity can vary. Semi-automatic and manual approaches are more subjective and dependent on the researcher's skills, but they may be necessary when working with special populations (e.g., infants[14]) whose signal shows less typical artifacts, difficult for automatic approaches to detect. Therefore, to increase transparency and reproducibility, the final report should include the list of epochs marked for rejection (or correction; see below). In addition, source code could be referenced in-text and be made publicly available for inspection upon publication, to provide additional information (e.g., on some parameter values) while decluttering the *Methods* section. An example of a planned automatic artifact rejection procedure could read as follows (modified from Schettino et al., 2020, p. 5): *FASTER v1.2.3b (Nolan et al., 2010) will be used for artifact identification and rejection: (i) the mean across channels will be computed for each epoch and, if amplitude range, variance, and channel deviation exceed z = ±3, the whole epoch will be removed; (ii) condition*

[14] Recently, though, automatic artifact rejection pipelines for infant EEG data have been developed, e.g., MADE (Debnath et al., 2020) and HAPPE (Gabard-Durnam et al., 2018).

averages with amplitude range, variance, channel deviation, and maximum EOG value exceeding z = ±3 will be removed; (iii) epochs containing more than 12 interpolated channels will be discarded. For exact parameter values, see the commented MATLAB script at https://www.example.com.[15]

In addition to artifact rejection procedures, there are a variety of techniques available to correct artifacts in order to reduce data loss. Methods for artifact correction include independent component analysis (ICA), regression-based methods, and wavelet-transforms (see Jiang et al., 2019). Each of these methods require different parameter values to be preregistered. For example, for ICA, at least the following information should be added (see, for instance, Keil et al., 2014): the method used to compute the ICA – e.g., fastICA (Hyvärinen, 1999) or infomax ICA (Bell and Sejnowski, 1997) –, the electrodes included, the number of computed components, the method by which artifactual components will be identified – e.g., using templates (Campos Viola et al., 2009) or manually –, and which type of artifactual components will be removed (e.g., only ocular components, ocular and heart beat components, etc.). For example, a preregistration for artifact correction using ICA could read: *We will use ICA to correct eye blinks and saccades. We will include all electrodes and use the fastICA algorithm (Hyvärinen, 1999) to compute 63 independent components. Components corresponding to blinks and saccades will be identified using the CORRMAP plugin for EEGLAB (Campos Viola et al., 2009).*

[15] Exclusion criteria based on data loss should also be included in the *Inclusion/Exclusion criteria* section of the preregistration. An example could read: *Participants with less than 70% remaining epochs will be excluded from data analysis.*

3.2.6 Baseline correction

A preregistration should also include whether or not the data will be baseline-corrected by setting the scalp distribution to zero during a preset period before the onset of the event of interest. In some cases (e.g., in language studies, where the baseline often contains part of the stimulus), not applying a baseline correction may be more appropriate. The preregistration should include the time window that is used as the baseline, and at which point in the analysis pipeline the baseline correction will be applied. For example: *We will apply baseline correction to the clean (i.e., after artifact rejection and correction), trial-averaged data, in a time window of -200 to 0 ms relative to the onset of the word.*

3.3 Statistical analysis

An example of ERP experiment may involve measurements from 64 electrodes and a sampling rate of 256 Hz, with a trial length of 1,000 ms after stimulus onset (e.g., Schindler et al., 2018). After averaging over trials, this would result in 16,384 data points for each participant and condition. For statistical analysis, this leads to a large number of potential comparisons, often referred to as the *multiple comparison problem* (MCP). Standard statistical correction procedures operating at the level of single electrode-time pairs would yield hyper-conservative results (increased *Type II* error); on the other hand, failing to correct for multiple comparisons can easily lead to spurious statistically significant results (increased *Type I* error). Therefore, statistical plans of ERP studies should always include a strategy on how to deal with the MCP. Several solutions are available (Luck, 2014, chap. 10; Luck and Gaspelin, 2017), including: (*i*) a priori definition of electrode sites and time windows based on previous studies; (*ii*) collapsed localizers, i.e., averaging all trials of all conditions of all

participants to identify electrode clusters and time windows, thereby avoiding condition-specific biases; and (*iii*) mass univariate statistics, i.e., computing statistical tests at each electrode and time point and applying appropriate multiple comparisons correction techniques (Fields and Kuperberg, 2020; Groppe et al., 2011).

In the following sections, we provide examples on how to preregister three common ERP analyses: (1) Analysis of Variance (ANOVA); (2) cluster-based permutation tests; and (3) Bayes factors.

3.3.1 Example 1: ANOVA

ANOVAs are a popular statistical technique to analyze ERP data: it is not uncommon to read published studies including several within- and between-subject factors for various spatial and temporal ROIs. However, a growing number of factors comes at a cost, namely an increase in false positive rate: up to 50% chance to find at least one false positive effect with 4 factors, and up to 100% chance with 8 factors (Luck and Gaspelin, 2017). Therefore, ERP researchers should carefully plan appropriate corrections for multiple testing not only for the follow-up tests to an ANOVA, but also as a function of the number of factors included in the ANOVA itself. Alternatively, researchers can limit the number of factors included in the ANOVA if they have a specific hypothesis about the spatial and temporal region of interest or only run planned (paired) contrasts on the relevant comparisons.

In the following example, researchers interested in the N400 ERP component (Kutas and Hillyard, 1980) identify the time window and region of interest *a priori*, and aim to test two hypotheses: (1) larger N400 effect for semantically incongruent compared to semantically congruent sentences (i.e., an incongruency effect); and (2) larger incongruency effect for native speakers compared to non-native speakers. In the *Analysis* section, the

preregistration could read: *We will analyze the amplitude*[16] *of the N400 by means of a mixed ANOVA with 2 factors: congruency (semantically congruent sentences vs. semantic violations; within-subject) and language experience (native speaker vs. non-native speaker; between-subject).* As outlined above, researchers should also consider correcting for multiple comparisons as a function of the number of factors in the ANOVA. In this example, three *p*-values will be computed: one for the main effect of congruency, one for the main effect of language experience, and one for their interaction. Researchers could preregister the correction in the following way: *The significance level for the main effect and interaction terms will be Bonferroni-corrected for the number of tests computed in the ANOVA: 0.05/3 = 0.0167.* Of course, the alpha level of the test should also be explicit (α = 0.05 in this example). If researchers also plan follow-up comparisons, they could add: *In case the interaction between congruency and language experience is statistically significant, we will compute two paired t-tests comparing congruent vs incongruent sentences, separately for native speakers and non-native speakers. For these two planned comparisons, we will use an alpha of 0.025 (Bonferroni-corrected for two planned tests with an uncorrected alpha of 0.05).*

3.3.2 Example 2: Cluster-based permutation tests

Cluster-based permutation tests (CBPT; Maris and Oostenveld, 2007) are another popular statistical approach to analyze EEG data. In the following example, the researchers have a hypothesis about a difference in ERP responses between two conditions, but not about specific electrodes and time points. The preregistration should include whether the test is

[16] It is also important to clearly specify the time window and electrode cluster from which the ERP component is scored. This information should be included in the section *Measured variables* of the preregistration protocol, e.g.: *We will analyze the mean amplitude value in a time window from 300 to 500 ms after stimulus onset, averaged across a cluster of centro-parietal electrodes (C3, Cz, C4, CP5, CP6, P3, Pz, P4, P7, and P8).*

one-tailed or two-tailed, within-subject, between-subject, or mixed, as well as the alpha level. In addition, there are several parameters that are more specific to the CBPT, such as the alpha at the cluster-level, the method for computing cluster statistics, the minimum number of electrodes that can form a cluster, how neighboring relations between electrodes will be computed, and the number of randomizations. For example, a preregistration using a CBPT could read: *To test within-subject differences between congruent and incongruent sentences, we will compute a two-tailed cluster-based permutation test using 'ft_statfun_indepsamplesT' in Fieldtrip (Oostenveld et al., 2011) with an alpha of 0.025 for each tail (i.e., the overall alpha is 0.05). The alpha at the cluster-level will be set at 0.05. Cluster statistics will be computed with a 'maxsum' approach and clusters will require a minimum of two neighboring electrodes. Neighboring electrodes will be defined via the 'triangulation' method implemented in Fieldtrip. Like on the test-level, the clusters will be tested with a two-tailed statistic. One thousand randomizations will be computed via Montecarlo method to estimate the p-value under the permutation distribution.*

3.3.3 Example 3: Bayes factors

Problems inherent in accepting the null hypothesis with classical frequentist procedures (e.g., Wagenmakers, 2007) and common misinterpretations of *p*-values (Colquhoun, 2017; Wasserstein and Lazar, 2016) are leading an increasing number of researchers to explore Bayesian approaches (Etz and Vandekerckhove, 2018; Kruschke and Liddell, 2017). Bayesian inference allows researchers to incorporate prior knowledge into statistical tests, quantify evidence in favor of the null hypothesis – thus discriminating between "absence of evidence" and "evidence of absence" (e.g., Keysers et al., 2020) –, and monitor the evidence as the data accumulate (Rouder, 2014; but see de Heide and Grünwald,

2020). In particular, Bayes factors – "*the extent to which the data sway our relative belief from one hypothesis to the other*" (Etz and Vandekerckhove, 2018, p. 10) – have gained considerable popularity, also thanks to the development of user-friendly software that facilitate their calculation (e.g., *JASP*; https://jasp-stats.org/).

Researchers planning to analyze their data using Bayes factors should clarify the software and procedure used for the estimation, a description of the prior specification (i.e., the type of distribution and its parameter values), and an assessment of the robustness of the results under different prior specifications (see also van Doorn et al., 2020). A preregistered description of planned comparisons using Bayes factors could read as follows (modified from Schindler et al., 2018, p. 10): *We will analyze the amplitude values of the N1 ERP component using Bayes Factors (BFs; Kass and Raftery, 1995). Two-tailed Bayesian t-tests (Rouder et al., 2009) will be calculated to estimate the degree of evidence in favor of a model assuming differences between conditions relative to a model assuming no differences. The null hypothesis will be specified as a point-null prior (Dirac distribution, standardized effect size* $\delta = 0$*), whereas the alternative hypothesis will be defined as a Jeffrey-Zellner-Siow (JZS) prior, a folded Cauchy distribution centered around* $\delta = 0$ *with scaling factors of* $r = 1$*,* $r = 0.707$*, and* $r = 0.5$*, to verify the robustness of the results as a function of changes in the prior (Schönbrodt et al., 2017). Participants will be included as random factors, and their variance considered nuisance. The threshold to identify the winning model is set at* $BF \geq 10$ *or* $BF \leq 0.1$*, typically considered "strong" evidence in favor of the model in the numerator or denominator, respectively (Kass and Raftery, 1995). BFs will be estimated via the R package BayesFactor v0.9.12-2 (Morey and Rouder, 2018) using Markov Chain Monte Carlo sampling (10,000 iterations).*

4. General Considerations

In *Section 2* we clarified how preregistration can mitigate some of the issues related to undisclosed analytic flexibility. In *Section 3* we provided guidelines and examples on how to preregister common preprocessing and statistical analysis steps in ERP studies. In *Section 4* we discuss several considerations that EEG researchers may want to take into account when critically evaluating whether to preregister their studies.

4.1 Preregistration vs. Registered Reports

Throughout this manuscript we have described *unreviewed* preregistrations (see van 't Veer and Giner-Sorolla, 2016), that is, the protocols uploaded on public repositories are not peer-reviewed. A preregistered study can still be rejected by scientific journals for a number of reasons – for example, lack of interest in "negative" (non-significant) findings (Fanelli, 2010), unprofessional peer-review (Gerwing et al., 2020), or submission in the "wrong" day of the week (Boja et al., 2018) –, thus limiting its discoverability (although the experimental protocol may still be publicly accessible). Conversely, *reviewed* preregistrations – commonly referred to as *Registered Reports* (Chambers and Tzavella, 2020) – are alternative article formats in which the study proposal is peer-reviewed and conditionally accepted for publication (in-principle acceptance, or *IPA*), provided that the original plan is followed and deviations are properly documented. Publication is thus independent from study outcome. Preliminary research has shown that this format seems to effectively mitigate publication bias and reduce the prevalence of suboptimal research practices (Scheel et al., 2020; see also Wiseman et al., 2019).[17] Therefore, we consider Registered Reports the state-of-the-art article

[17] Due to heterogeneous journal policies, IPA protocols may not always be publicly available or easily verifiable (Hardwicke and Ioannidis, 2018), which sometimes makes it difficult for readers to compare registered plans and published papers. Recent developments have tackled the lack of transparency and standardized protocol

format for confirmatory research and recommend them over preregistrations. At the time of writing, more than 250 journals[18] from various scholarly disciplines offer Registered Reports alongside traditional submissions, including *Psychophysiology* (Keil et al., 2020) and the *International Journal of Psychophysiology* (Larson, 2016). Please note that, while the current manuscript focuses on preregistrations, our recommendations also hold for Registered Reports.

Despite these desirable properties, researchers should take into account the relatively strict submission criteria (e.g., expected statistical power of 90% or higher[19]) as well as the time necessary to review the study plan (typically 2-4 months; Chambers, 2020), during which the project cannot start. For these reasons, preregistration can be seen as an easier, entry-level practice that is advantageous in itself and helps researchers familiarize with the steps required for a future (recommended) Registered Report submission.

4.2 Perceived Disadvantages of Preregistration

Crafting a comprehensive preregistration protocol requires time. A recent survey (Toth et al., 2020) revealed that respondents with previous experience in preregistration invested, on average, around 4 hours to create the initial draft. Our experience suggests that this could be considered a lower bound: the multidimensional nature of EEG data, coupled with the high level of specificity recommended to effectively avoid selective reporting, requires documenting a large number of preprocessing and analysis steps (see *Section 3*), planned sample size (with *a priori* power analysis), sampling strategy, inclusion and exclusion criteria, and more. However, most of the information included in a comprehensive preregistration is also required in the final manuscript, not only to facilitate communication

registration by updating recommended editorial policy templates (Chambers and Mellor, 2018), but we advise to read the target journal's specific guidelines before submission.

[18] The updated list of journals offering Registered Reports can be found at https://cos.io/rr/.

[19] As an example, see *Registered Reports Submission Guidelines* at *Cortex* (https://tinyurl.com/RR-Cortex).

between the authors and other relevant parties (editors, reviewers, and readers) but also to ensure that the methods leading to the conclusions advertised in the paper are reproducible. For instance, publication guidelines for M/EEG studies (Gross et al., 2013; Keil et al., 2014; Pernet et al., 2018) emphasize the need to describe equipment, study materials, preprocessing steps, dependent variables, and analysis pipelines, and provide a checklist that authors can consult while writing the manuscript (e.g., Keil et al., 2014, sec. Appendix). Here we propose to anticipate this time investment, with the advantage that carefully thinking about these methodological details before data collection may lead to improvements in the study design when still useful.

Some researchers might also be worried that the time invested in writing the preregistration would be wasted if results do not pan out as expected and, consequently, the final manuscript would be more difficult to publish. Indeed, the pressure to publish novel, groundbreaking, positive results is pervasive in many disciplines (Fanelli, 2012, 2010; Ioannidis et al., 2014; Jennings and Van Horn, 2012; Nissen et al., 2016; Scargle, 2000). Nonetheless, many respectable academic journals accept manuscripts with non-significant findings if the methodology is robust, with the aim to mitigate the pervasive problem of publication bias.[20]

Unforeseen circumstances – e.g., problems recruiting the planned number of participants due to a pandemic – may require reasonable deviations from the original preregistered plan. This is acceptable as long as it is transparently documented in the published paper. Regrettably, recent evidence shows that undisclosed protocol deviations are common. An analysis of preregistered studies in the journal *Psychological Science* (Claesen

[20] See, for instance, two recent *Nature* editorials (2020; 2017), the *PLOS ONE* article collection *Missing Pieces* (2015), the editorial by Munafò and Neill (2016) in the *Journal of Psychopharmacology*, as well as the submission guidelines of *Meta-Psychology* (https://open.lnu.se/index.php/metapsychology/about) and *Royal Society Open Science* (https://royalsocietypublishing.org/rsos/for-authors).

et al., 2019) showed that none of them had perfectly anticipated every step of the research project: differences between preregistration protocols and final manuscripts were observed, for instance, in sample size, exclusion criteria, and statistical models, with only one study transparently reporting all discrepancies. Partial or lack of disclosure of deviations from pre-study plans is a well-known problem not only in social sciences (e.g., Franco et al., 2016) but also in clinical trials (Goldacre et al., 2019), for which registration is compulsory or strongly supported in many countries (see the *World Health Organization Registry Network*; https://www.who.int/ictrp/network/en/). We emphatically recommend to clearly report any deviations from the preregistered plan, preferably in a separate section in the main manuscript or in *Supplementary Materials*. A useful checklist can be found on the OSF (https://osf.io/yrvcg/).

4.3 Preregistration Is Not a Silver Bullet

As argued *in Section 2.1*, preregistration can strengthen the evidential value of studies by clarifying the distinction between hypothesis- and data-driven analysis, preventing HARKing, disclosing selective outcome reporting, and increasing the number of publications with non-significant findings (see also Nosek et al., 2019). However, adopted in isolation, preregistration may *not* necessarily prompt researchers to carefully examine whether their chosen statistical models are appropriate for the experimental question (Guest and Martin, 2020; Szollosi et al., 2020), improve the link between theories and their mathematical representations (Szollosi and Donkin, 2019b), or develop more precise, consistent, and "hard-to-vary" theories altogether (Szollosi and Donkin, 2019a). In other words, preregistration in itself does not necessarily improve the quality of the research. However, it does increase the ability of the tests within a study to falsify predictions, i.e., their *severity* (e.g., Lakens, 2019). Furthermore, it offers a window into the research workflow, an often

messy and non-linear process that is far from the flawless stories recounted in academic papers. Accepting these imperfections may promote a work culture that normalizes errors, acknowledges the depth of domain-specific knowledge, and fosters intra- and interdisciplinary collaborations (see also Nosek et al., 2012).

Finally, just like other research practices, preregistration can in principle be used unethically (Yamada, 2018). For example, one could preregister a large number of similar experiments and keep them under embargo. Whenever one of the studies turns out to be "successful" (i.e., statistically significant), the resulting paper would only refer to the corresponding preregistration and all the other "unsuccessful" ones could be withdrawn. However, metadata and a justification for each withdrawal would still be publicly available and therefore raise suspicion. Another unethical practice has been termed "preregistering after the results are known" (PARKing; Yamada, 2018), i.e., drafting and publishing the preregistration of a study that has already been completed and whose results conveniently fit within the narrative of the "preregistered" document. Complementary open science practices, e.g., data and code sharing, can effectively mitigate this risk. Journal editors and reviewers are invited to carefully compare the preregistered document with the manuscript during submission and evaluation, as well as request raw data and analysis code by default (Morey et al., 2016). Having said that, we prefer to think of fellow researchers in more optimistic terms, motivated by higher goals than simply publishing as much as possible (although we might, admittedly, be hopelessly naive; Chapman et al., 2019; DeDeo, 2020). We also point out that premeditated approaches to exploit the vulnerabilities of a system can hardly be reconciled with claims that mistakes were made in good faith due to ignorance or procedural complexity: researchers engaging in such behaviors (if proven) would consciously commit fraud, and responsible institutions should be contacted and deliver appropriate sanctions.

5. Conclusion

The adoption of a new technique can be met with resistance, particularly if the benefits are unclear, the amount of work is perceived as too onerous, and training and guidance are lacking. In this paper, we argue that preregistering EEG projects can effectively limit flexibility in data preprocessing and analysis, thereby decreasing the risk of false positive findings and inflated effect sizes which ultimately impact study replicability and the reliability of accumulated knowledge. The time spent writing a preregistration is saved at later stages, because the information included in a comprehensive protocol is required at the time of publication. Ready-made templates can serve as useful guidelines and facilitate the implementation of this practice in the research workflow. Combined with other open science practices – e.g., sharing study protocols, materials, raw data, and analysis code –, preregistration increases transparency in the research process and trustworthiness of the scholarly products not only for academic peers, but also other stakeholders in society (Jamieson et al., 2019).

Contributions

AS supervised the project. All authors wrote the initial draft, reviewed, edited, and approved the final version of the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The authors would like to thank Dr. Matt Craddock for useful comments on an earlier version of the manuscript.

Competing Interests

MP is funded by the Max Planck Society and the Berlin School of Mind and Brain. GHG is funded by the Max Planck Society and the Einstein Center for Neurosciences Berlin. AS is employed at Erasmus University Rotterdam as Coordinator Open Science. Funders and employers had no role in project design, literature review, writing of the paper, and/or decision to submit for publication.

References

Alday, P.M., 2019. How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. Psychophysiology 56, e13451. https://doi.org/10.1111/psyp.13451

Algermissen, J., Bartlett, J., Gau, R., Heunis, S., Klapwijk, E., Mazor, M., Paul, M., Schettino, A., Mehler, D.M.A., 2019. A Fruitful Rendezvous at SIPS: Neuroimagers Meet Study Preregistration Advocates [WWW Document]. Organ. Hum. Brain Mapp. URL http://www.ohbmbrainmappingblog.com/1/post/2019/10/a-fruitful-rendevous-at-sips-neuroimagers-meet-study-preregistration-advocates.html (accessed 10.29.20).

Allen, C., Mehler, D.M.A., 2019. Open science challenges, benefits and tips in early career and beyond. PLOS Biol. 17, e3000246. https://doi.org/10.1371/journal.pbio.3000246

Anand, V., Scales, D.C., Parshuram, C.S., Kavanagh, B.P., 2014. Registration and design alterations of clinical trials in critical care: a cross-sectional observational study. Intensive Care Med. 40, 700–722. https://doi.org/10.1007/s00134-014-3250-7

Babbage, C., 1830. Reflections on the decline of science in England, and on some of its causes. B. Fellowes [etc.], London.

Bakker, M., Veldkamp, C.L.S., van Assen, M.A.L.M., Crompvoets, E.A.V., Ong, H.H., Nosek, B.A., Soderberg, C.K., Mellor, D.T., Wicherts, J.M., 2018. Ensuring the quality and specificity of preregistrations (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/cdgyh

Baldwin, S.A., 2017. Improving the rigor of psychophysiology research. Int. J. Psychophysiol. 111, 5–16. https://doi.org/10.1016/j.ijpsycho.2016.04.006

Bastos, A.M., Schoffelen, J.-M., 2016. A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls. Front. Syst. Neurosci. 9. https://doi.org/10.3389/fnsys.2015.00175

Bell, A.J., Sejnowski, T.J., 1997. The “independent components” of natural scenes are edge filters. Vision Res. 37, 3327–3338. https://doi.org/10/dcrkkm

Bem, D.J., 2011. Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. J. Pers. Soc. Psychol. 100, 407–425. https://doi.org/10.1037/a0021524

Bigdely-Shamlo, N., Touryan, J., Ojeda, A., Kothe, C., Mullen, T., Robbins, K., 2020a. Automated EEG mega-analysis I: Spectral and amplitude characteristics across studies. NeuroImage 207, 116361. https://doi.org/10.1016/j.neuroimage.2019.116361

Bigdely-Shamlo, N., Touryan, J., Ojeda, A., Kothe, C., Mullen, T., Robbins, K., 2020b. Automated EEG mega-analysis II: Cognitive aspects of event related features. NeuroImage 207, 116054. https://doi.org/10.1016/j.neuroimage.2019.116054

Board of Governors of the Federal Reserve System, 2015. Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say “Usually Not.” Finance Econ. Discuss. Ser. 2015, 1–26. https://doi.org/10.17016/FEDS.2015.083

Boja, C.E., Herţeliu, C., Dârdală, M., Ileanu, B.V., 2018. Day of the week submission effect for accepted papers in Physica A, PLOS ONE, Nature and Cell. Scientometrics 117, 887–918. https://doi.org/10.1007/s11192-018-2911-7

Botvinik-Nezer, R., Holzmeister, F., Camerer, C.F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J.A., Adcock, R.A., Avesani, P., Baczkowski,

B.M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., Benoit, R.G., Berkers, R.M.W.J., Bhanji, J.P., Biswal, B.B., Bobadilla-Suarez, S., Bortolini, T., Bottenhorn, K.L., Bowring, A., Braem, S., Brooks, H.R., Brudner, E.G., Calderon, C.B., Camilleri, J.A., Castrellon, J.J., Cecchetti, L., Cieslik, E.C., Cole, Z.J., Collignon, O., Cox, R.W., Cunningham, W.A., Czoschke, S., Dadi, K., Davis, C.P., De Luca, A., Delgado, M.R., Demetriou, L., Dennison, J.B., Di, X., Dickie, E.W., Dobryakova, E., Donnat, C.L., Dukart, J., Duncan, N.W., Durnez, J., Eed, A., Eickhoff, S.B., Erhart, A., Fontanesi, L., Fricke, G.M., Fu, S., Galván, A., Gau, R., Genon, S., Glatard, T., Glerean, E., Goeman, J.J., Golowin, S.A.E., González-García, C., Gorgolewski, K.J., Grady, C.L., Green, M.A., Guassi Moreira, J.F., Guest, O., Hakimi, S., Hamilton, J.P., Hancock, R., Handjaras, G., Harry, B.B., Hawco, C., Herholz, P., Herman, G., Heunis, S., Hoffstaedter, F., Hogeveen, J., Holmes, S., Hu, C.-P., Huettel, S.A., Hughes, M.E., Iacovella, V., Iordan, A.D., Isager, P.M., Isik, A.I., Jahn, A., Johnson, M.R., Johnstone, T., Joseph, M.J.E., Juliano, A.C., Kable, J.W., Kassinopoulos, M., Koba, C., Kong, X.-Z., Koscik, T.R., Kucukboyaci, N.E., Kuhl, B.A., Kupek, S., Laird, A.R., Lamm, C., Langner, R., Lauharatanahirun, N., Lee, H., Lee, S., Leemans, A., Leo, A., Lesage, E., Li, F., Li, M.Y.C., Lim, P.C., Lintz, E.N., Liphardt, S.W., Losecaat Vermeer, A.B., Love, B.C., Mack, M.L., Malpica, N., Marins, T., Maumet, C., McDonald, K., McGuire, J.T., Melero, H., Méndez Leal, A.S., Meyer, B., Meyer, K.N., Mihai, G., Mitsis, G.D., Moll, J., Nielson, D.M., Nilsonne, G., Notter, M.P., Olivetti, E., Onicas, A.I., Papale, P., Patil, K.R., Peelle, J.E., Pérez, A., Pischedda, D., Poline, J.-B., Prystauka, Y., Ray, S., Reuter-Lorenz, P.A., Reynolds, R.C., Ricciardi, E., Rieck, J.R., Rodriguez-Thompson, A.M., Romyn, A., Salo, T., Samanez-Larkin, G.R., Sanz-Morales, E., Schlichting, M.L., Schultz, D.H., Shen, Q., Sheridan, M.A., Silvers, J.A., Skagerlund, K., Smith, A., Smith, D.V., Sokol-Hessner, P., Steinkamp, S.R., Tashjian, S.M., Thirion, B., Thorp, J.N., Tinghög, G., Tisdall, L., Tompson, S.H., Toro-Serey, C., Torre Tresols, J.J., Tozzi, L., Truong, V., Turella, L., van ‘t Veer, A.E., Verguts, T., Vettel, J.M., Vijayarajah, S., Vo, K., Wall, M.B., Weeda, W.D., Weis, S., White, D.J., Wisniewski, D., Xifra-Porxas, A., Yearling, E.A., Yoon, S., Yuan, R., Yuen, K.S.L., Zhang, L., Zhang, X., Zosky, J.E., Nichols, T.E., Poldrack, R.A., Schonberg, T., 2020. Variability in the analysis of a single neuroimaging dataset by many teams. Nature 1–7. https://doi.org/10.1038/s41586-020-2314-9

Bourgeois, F.T., Murthy, S., Mandl, K.D., 2010. Outcome reporting among drug trials registered in ClinicalTrials.gov. Ann. Intern. Med. 153, 158–166. https://doi.org/10/ghdcwb

Bowring, A., Maumet, C., Nichols, T.E., 2019. Exploring the impact of analysis software on task fMRI results. Hum. Brain Mapp. 40, 3362–3384. https://doi.org/10.1002/hbm.24603

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365–376. https://doi.org/10.1038/nrn3475

Camerer, C.F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., Wu, H., 2016. Evaluating replicability of laboratory experiments in economics. Science 351, 1433–1436. https://doi.org/10.1126/science.aaf0918

Campos Viola, F., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., Debener, S., 2009.

Semi-automatic identification of independent components representing EEG artifact. Clin. Neurophysiol. 120, 868–877. https://doi.org/10/czzxhw

Carp, J., 2012a. The secret lives of experiments: Methods reporting in the fMRI literature. NeuroImage 63, 289–300. https://doi.org/10.1016/j.neuroimage.2012.07.004

Carp, J., 2012b. On the Plurality of (Methodological) Worlds: Estimating the Analytic Flexibility of fMRI Experiments. Front. Neurosci. 6. https://doi.org/10.3389/fnins.2012.00149

Chalmers, I., Bracken, M.B., Djulbegovic, B., Garattini, S., Grant, J., Gülmezoglu, A.M., Howells, D.W., Ioannidis, J.P.A., Oliver, S., 2014. How to increase value and reduce waste when research priorities are set. The Lancet 383, 156–165. https://doi.org/10.1016/S0140-6736(13)62229-1

Chambers, C.. D., Tzavella, L., 2020. Registered Reports: Past, Present and Future (preprint). MetaArXiv. https://doi.org/10.31222/osf.io/43298

Chambers, C.D., 2020. Why we need preregistration of research: Introducing “Registered Reports.”

Chambers, C.D., Mellor, D.T., 2018. Protocol transparency is vital for registered reports. Nat. Hum. Behav. 2, 791–792. https://doi.org/10.1038/s41562-018-0449-6

Chang, C.-Y., Hsu, S.-H., Pion-Tonachini, L., Jung, T.-P., 2020. Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings. IEEE Trans. Biomed. Eng. 67, 1114–1121. https://doi.org/10.1109/TBME.2019.2930186

Chapman, C.A., Bicca-Marques, J.C., Calvignac-Spencer, S., Fan, P., Fashing, P.J., Gogarten, J., Guo, S., Hemingway, C.A., Leendertz, F., Li, B., Matsuda, I., Hou, R., Serio-Silva, J.C., Chr. Stenseth, N., 2019. Games academics play and their consequences: how authorship, *h* -index and journal impact factors are shaping the future of academia. Proc. R. Soc. B Biol. Sci. 286, 20192047. https://doi.org/10/ggd7dk

Claesen, A., Gomes, S.L.B.T., Tuerlinckx, F., Vanpaemel, W., 2019. Preregistration: Comparing Dream to Reality (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/d8wex

Clayson, P.E., Carbine, K.A., Baldwin, S.A., Larson, M.J., 2019. Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability. Psychophysiology 56, e13437. https://doi.org/10.1111/psyp.13437

Clayson, P.E., Miller, G.A., 2017. Psychometric considerations in the measurement of event-related brain potentials: Guidelines for measurement and reporting. Int. J. Psychophysiol. 111, 57–67. https://doi.org/10.1016/j.ijpsycho.2016.09.005

Cohen, M.X., 2014. Analyzing Neural Time Series Data: Theory and Practice, 1 edition. ed. The MIT Press, Cambridge, Massachusetts.

Colquhoun, D., 2017. The reproducibility of research and the misinterpretation of *p* -values. R. Soc. Open Sci. 4, 171085. https://doi.org/10/gd6dq2

Cumming, G., 2014. The New Statistics: Why and How. Psychol. Sci. 25, 7–29. https://doi.org/10.1177/0956797613504966

de Heide, R., Grünwald, P.D., 2020. Why optional stopping is a problem for Bayesians. ArXiv170808278 Math Stat.

Debnath, R., Buzzell, G.A., Morales, S., Bowers, M.E., Leach, S.C., Fox, N.A., 2020. The Maryland analysis of developmental EEG (MADE) pipeline. Psychophysiology 57. https://doi.org/10/ghdh8p

DeDeo, S., 2020. When Science is a Game. ArXiv200605994 Phys.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21. https://doi.org/10/bqr2f2

Delorme, A., Sejnowski, T., Makeig, S., 2007. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. NeuroImage 34, 1443–1449. https://doi.org/10/dvrjq7

Depaoli, S., van de Schoot, R., 2017. Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. Psychol. Methods 22, 240–261. https://doi.org/10/f99qn4

Devezer, B., Navarro, D.J., Vandekerckhove, J., Buzbas, E.O., 2020. The case for formal methodology in scientific reform (preprint). Scientific Communication and Education. https://doi.org/10.1101/2020.04.26.048306

Dirnagl, U., 2020. Preregistration of exploratory research: Learning from the golden age of discovery. PLOS Biol. 18, e3000690. https://doi.org/10/gg9fvb

Dong, L., Liu, X., Zhao, L., Lai, Y., Gong, D., Liu, T., Yao, D., 2019. A Comparative Study of Different EEG Reference Choices for Event-Related Potentials Extracted by Independent Component Analysis. Front. Neurosci. 13. https://doi.org/10.3389/fnins.2019.01068

Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Natl. Acad. Sci. 113, 7900–7905. https://doi.org/10.1073/pnas.1602413113

Enders, C.K., 2010. Applied missing data analysis, Applied missing data analysis. Guilford Press, New York, NY, US.

Etz, A., Vandekerckhove, J., 2018. Introduction to Bayesian Inference for Psychology. Psychon. Bull. Rev. 25, 5–34. https://doi.org/10/gc9pqc

Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L., 1991. Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Electroencephalogr. Clin. Neurophysiol. 78, 447–455. https://doi.org/10.1016/0013-4694(91)90062-9

Fanelli, D., 2012. Negative results are disappearing from most disciplines and countries. Scientometrics 90, 891–904. https://doi.org/10/bq24dv

Fanelli, D., 2010. “Positive” Results Increase Down the Hierarchy of the Sciences. PLoS ONE 5, e10068. https://doi.org/10/d3bcr7

Field, A.P., Miles, J.N.V., Field, Z.C., 2012. Discovering statistics using R: and sex and drugs and rock “n” roll. Sage Publications Ltd., London.

Fields, E.C., Kuperberg, G.R., 2020. Having your cake and eating it too: Flexibility and power with mass univariate statistics for ERP data. Psychophysiology 57, e13468. https://doi.org/10.1111/psyp.13468

Fletcher, E.M., Kussmaul, C.L., Mangun, G.R., 1996. Estimation of interpolation errors in scalp topographic mapping. Electroencephalogr. Clin. Neurophysiol. 98, 422–434. https://doi.org/10.1016/0013-4694(96)95135-4

Franco, A., Malhotra, N., Simonovits, G., 2016. Underreporting in Psychology Experiments: Evidence From a Study Registry. Soc. Psychol. Personal. Sci. 7, 8–12. https://doi.org/10/f75sr6

Frömer, R., Maier, M., Abdel Rahman, R., 2018. Group-Level EEG-Processing Pipeline for Flexible Single Trial-Based Analyses Including Linear Mixed Models. Front. Neurosci. 12, 48. https://doi.org/10/gc3z35

Gabard-Durnam, L.J., Mendez Leal, A.S., Wilkinson, C.L., Levin, A.R., 2018. The Harvard

Automated Processing Pipeline for Electroencephalography (HAPPE): Standardized Processing Software for Developmental and High-Artifact Data. Front. Neurosci. 12, 97. https://doi.org/10/gc6c3z

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., Donchin, E., 1993. A Neural System for Error Detection and Compensation. Psychol. Sci. 4, 385–390.

Gelman, A., Loken, E., 2013. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. Columbia Stat.

Gerwing, T.G., Allen Gerwing, A.M., Avery-Gomm, S., Choi, C.-Y., Clements, J.C., Rash, J.A., 2020. Quantifying professionalism in peer review. Res. Integr. Peer Rev. 5, 9. https://doi.org/10.1186/s41073-020-00096-x

Gigerenzer, G., 2004. Mindless statistics. J. Socio-Econ., Statistical Significance 33, 587–606. https://doi.org/10.1016/j.socec.2004.09.033

Goldacre, B., Drysdale, H., Dale, A., Milosevic, I., Slade, E., Hartley, P., Marston, C., Powell-Smith, A., Heneghan, C., Mahtani, K.R., 2019. COMPare: a prospective cohort study correcting and monitoring 58 misreported trials in real time. Trials 20, 118. https://doi.org/10/gfvpv9

Grinde, K.E., Arbet, J., Green, A., O’Connell, M., Valcarcel, A., Westra, J., Tintle, N., 2017. Illustrating, Quantifying, and Correcting for Bias in Post-hoc Analysis of Gene-Based Rare Variant Tests of Association. Front. Genet. 8, 117. https://doi.org/10.3389/fgene.2017.00117

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. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Gross, J., Baillet, S., Barnes, G.R., Henson, R.N., Hillebrand, A., Jensen, O., Jerbi, K., Litvak, V., Maess, B., Oostenveld, R., Parkkonen, L., Taylor, J.R., van Wassenhove, V., Wibral, M., Schoffelen, J.-M., 2013. Good practice for conducting and reporting MEG research. NeuroImage 65, 349–363. https://doi.org/10.1016/j.neuroimage.2012.10.001

Guest, O., Martin, A.E., 2020. How computational modeling can force theory building in psychological science (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/rybh9

Hajcak, G., Klawohn, J., Meyer, A., 2019. The Utility of Event-Related Potentials in Clinical Psychology. Annu. Rev. Clin. Psychol. 15, 71–95. https://doi.org/10.1146/annurev-clinpsy-050718-095457

Hannink, G., Gooszen, H.G., Rovers, M.M., 2013. Comparison of registered and published primary outcomes in randomized clinical trials of surgical interventions. Ann. Surg. 257, 818–823. https://doi.org/10.1097/SLA.0b013e3182864fa3

Hansen, P., Kringelbach, M., Salmelin, R. (Eds.), 2010. MEG: An Introduction to Methods, 1 edition. ed. Oxford University Press, New York.

Harder, R.L., Desmarais, R.N., 1972. Interpolation using surface splines. J. Aircr. 9, 189–191. https://doi.org/10.2514/3.44330

Hardwicke, T.E., Ioannidis, J.P.A., 2018. Mapping the universe of registered reports. Nat. Hum. Behav. 2, 793–796. https://doi.org/10.1038/s41562-018-0444-y

Haven, T.L., Van Grootel, L., 2019. Preregistering qualitative research. Account. Res. 26, 229–244. https://doi.org/10/gfv55q

Hawkins, D., 1980. Identification of Outliers, Monographs on Statistics and Applied Probability. Springer Netherlands. https://doi.org/10.1007/978-94-015-3994-4

Helfrich, R.F., Knight, R.T., 2019. Cognitive neurophysiology: Event-related potentials.

Handb. Clin. Neurol. 160, 543–558. https://doi.org/10.1016/B978-0-444-64032-1.00036-9

Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. 10, 626–634. https://doi.org/10/ftk87t

In praise of replication studies and null results, 2020. . Nature 578, 489–490. https://doi.org/10.1038/d41586-020-00530-6

Ioannidis, J.P.A., 2008. Why most discovered true associations are inflated. Epidemiol. Camb. Mass 19, 640–648. https://doi.org/10.1097/EDE.0b013e31818131e7

Ioannidis, J.P.A., 2005. Why most published research findings are false. PLoS Med. 2, e124.

Ioannidis, J.P.A., Munafò, M.R., Fusar-Poli, P., Nosek, B.A., David, S.P., 2014. Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. Trends Cogn. Sci. 18, 235–241. https://doi.org/10.1016/j.tics.2014.02.010

Jamieson, K.H., McNutt, M., Kiermer, V., Sever, R., 2019. Signaling the trustworthiness of science. Proc. Natl. Acad. Sci. 116, 19231–19236. https://doi.org/10/gf85w6

Jennings, R.G., Van Horn, J.D., 2012. Publication bias in neuroimaging research: implications for meta-analyses. Neuroinformatics 10, 67–80. https://doi.org/10.1007/s12021-011-9125-y

Jiang, X., Bian, G.-B., Tian, Z., 2019. Removal of Artifacts from EEG Signals: A Review. Sensors 19, 987. https://doi.org/10.3390/s19050987

John, L.K., Loewenstein, G., Prelec, D., 2012. Measuring the prevalence of questionable research practices with incentives for truth telling. Psychol. Sci. 23, 524–532. https://doi.org/10.1177/0956797611430953

Jones, P.M., Chow, J.T.Y., Arango, M.F., Fridfinnson, J.A., Gai, N., Lam, K., Turkstra, T.P., 2017. Comparison of Registered and Reported Outcomes in Randomized Clinical Trials Published in Anesthesiology Journals. Anesth. Analg. 125, 1292–1300. https://doi.org/10.1213/ANE.0000000000002272

Judd, C.M., McClelland, G.H., Ryan, C.S., 2017a. Data Analysis: A Model Comparison Approach To Regression, ANOVA, and Beyond, 3 edition. ed. Routledge, New York.

Judd, C.M., Westfall, J., Kenny, D.A., 2017b. Experiments with More Than One Random Factor: Designs, Analytic Models, and Statistical Power. Annu. Rev. Psychol. 68, 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

Kaplan, R.M., Irvin, V.L., 2015. Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time. PLoS ONE 10. https://doi.org/10.1371/journal.pone.0132382

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. https://doi.org/10.1111/j.1469-8986.2010.01009.x

Kass, R.E., Raftery, A.E., 1995. Bayes Factors. J. Am. Stat. Assoc. 90, 773–795. https://doi.org/10/gdnbw3

Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E.S., Luck, S.J., Luu, P., Miller, G.A., Yee, C.M., 2014. Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. Psychophysiology 51, 1–21. https://doi.org/10.1111/psyp.12147

Keil, A., Gatzke-Kopp, L.M., Horváth, J., Jennings, J.R., Fabiani, M., 2020. A registered report format for *Psychophysiology*. Psychophysiology. https://doi.org/10.1111/psyp.13663

Kerr, N.L., 1998. HARKing: hypothesizing after the results are known. Personal. Soc.

Psychol. Rev. Off. J. Soc. Personal. Soc. Psychol. Inc 2, 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Keysers, C., Gazzola, V., Wagenmakers, E.-J., 2020. Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. Nat. Neurosci. 23, 788–799. https://doi.org/10/gg3tz7

Kida, T., Tanaka, E., Kakigi, R., 2016. Multi-Dimensional Dynamics of Human Electromagnetic Brain Activity. Front. Hum. Neurosci. 9. https://doi.org/10.3389/fnhum.2015.00713

Kidwell, M.C., Lazarević, L.B., Baranski, E., Hardwicke, T.E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T.M., Fiedler, S., Nosek, B.A., 2016. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. PLOS Biol. 14, e1002456. https://doi.org/10.1371/journal.pbio.1002456

Klawohn, J., Meyer, A., Weinberg, A., Hajcak, G., 2020. Methodological choices in event-related potential (ERP) research and their impact on internal consistency reliability and individual differences: An examination of the error-related negativity (ERN) and anxiety. J. Abnorm. Psychol. 129, 29–37. https://doi.org/10.1037/abn0000458

Kraemer, H.C., Mintz, J., Noda, A., Tinklenberg, J., Yesavage, J.A., 2006. Caution Regarding the Use of Pilot Studies to Guide Power Calculations for Study Proposals. Arch. Gen. Psychiatry 63, 484. https://doi.org/10/bqq4s2

Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., Baker, C.I., 2009. Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. 12, 535–540. https://doi.org/10.1038/nn.2303

Kruschke, J.K., 2014. Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan, 2nd edition. ed. Academic Press, Boston.

Kruschke, J.K., Liddell, T.M., 2017. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychon. Bull. Rev. 25, 178–206. https://doi.org/10.3758/s13423-016-1221-4

Kunda, Z., 1990. The case for motivated reasoning. Psychol. Bull. 108, 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205. https://doi.org/10/ffh9zj

Lakens, D., 2019. The Value of Preregistration for Psychological Science: A Conceptual Analysis (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/jbh4w

Lakens, D., Albers, C.J., 2017. When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/b7z4q

Lamberink, H.J., Otte, W.M., Sinke, M.R.T., Lakens, D., Glasziou, P.P., Tijdink, J.K., Vinkers, C.H., 2018. Statistical power of clinical trials increased while effect size remained stable: an empirical analysis of 136,212 clinical trials between 1975 and 2014. J. Clin. Epidemiol. 102, 123–128. https://doi.org/10.1016/j.jclinepi.2018.06.014

Lane, D.M., Dunlap, W.P., 1978. Estimating effect size: Bias resulting from the significance criterion in editorial decisions. Br. J. Math. Stat. Psychol. 31, 107–112. https://doi.org/10.1111/j.2044-8317.1978.tb00578.x

Larson, M.J., 2016. Commitment to Cutting-Edge Research with Rigor and Replication in Psychophysiological Science by Michael J. Larson. Int. J. Psychophysiol. 102, ix–x. https://doi.org/10.1016/S0167-8760(16)30038-1

Larson, M.J., Carbine, K.A., 2017. Sample size calculations in human electrophysiology (EEG and ERP) studies: A systematic review and recommendations for increased rigor. Int. J. Psychophysiol., Rigor and Replication: Towards Improved Best Practices in Psychophysiological Research 111, 33–41. https://doi.org/10.1016/j.ijpsycho.2016.06.015

Law, S.K., Nunez, P.L., Wijesinghe, R.S., 1993. High-resolution EEG using spline generated surface Laplacians on spherical and ellipsoidal surfaces. IEEE Trans. Biomed. Eng. 40, 145–153. https://doi.org/10.1109/10.212068

Leon, A.C., Davis, L.L., Kraemer, H.C., 2011. The role and interpretation of pilot studies in clinical research. J. Psychiatr. Res. 45, 626–629. https://doi.org/10/cmhgtj

Levelt, W.J.M., Drenth, P., Noort, E. (Eds.), 2012. Flawed science: The fraudulent research practices of social psychologist Diederik Stapel. Tilburg: Commissioned by the Tilburg University, University of Amsterdam and the University of Groningen.

Liu, Q., Balsters, J.H., Baechinger, M., van der Groen, O., Wenderoth, N., Mantini, D., 2015. Estimating a neutral reference for electroencephalographic recordings: the importance of using a high-density montage and a realistic head model. J. Neural Eng. 12, 056012. https://doi.org/10.1088/1741-2560/12/5/056012

Luck, S.J., 2014. An Introduction to the Event-Related Potential Technique, 2nd edition. ed. Bradford Books, Cambridge, Massachusetts.

Luck, S.J., Gaspelin, N., 2017. How to get statistically significant effects in any ERP experiment (and why you shouldn't). Psychophysiology 54, 146–157. https://doi.org/10.1111/psyp.12639

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190. https://doi.org/10/dt923p

McElreath, R., 2018. Statistical Rethinking: A Bayesian Course with Examples in R and Stan. Chapman and Hall/CRC. https://doi.org/10.1201/9781315372495

McKiernan, E.C., Bourne, P.E., Brown, C.T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B.A., Ram, K., Soderberg, C.K., Spies, J.R., Thaney, K., Updegrove, A., Woo, K.H., Yarkoni, T., 2016. How open science helps researchers succeed. eLife 5, e16800. https://doi.org/10.7554/eLife.16800

McPhetres, J., 2020. What should a preregistration contain? (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/cj5mh

Meehl, P.E., 1967. Theory-Testing in Psychology and Physics: A Methodological Paradox. Philos. Sci. 34, 103–115. https://doi.org/10/d7nnf3

Mensen, A., Khatami, R., 2013. Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics. NeuroImage 67, 111–118. https://doi.org/10/ghffpj

Mertens, G., Krypotos, A.-M., 2019. Preregistration of Analyses of Preexisting Data. Psychol. Belg. 59, 338–352. https://doi.org/10.5334/pb.493

Michel, C.M., He, B., 2019. EEG source localization. Handb. Clin. Neurol. 160, 85–101. https://doi.org/10.1016/B978-0-444-64032-1.00006-0

Michel, C.M., Koenig, T., 2018. EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. NeuroImage 180, 577–593. https://doi.org/10.1016/j.neuroimage.2017.11.062

Morey, R.D., Chambers, C.D., Etchells, P.J., Harris, C.R., Hoekstra, R., Lakens, D., Lewandowsky, S., Morey, C.C., Newman, D.P., Schönbrodt, F.D., Vanpaemel, W., Wagenmakers, E.-J., Zwaan, R.A., 2016. The Peer Reviewers' Openness Initiative: incentivizing open research practices through peer review. R. Soc. Open Sci. 3,

150547. https://doi.org/10/gf4293
Morey, R.D., Rouder, J.N., 2018. BayesFactor: Computation of bayes factors for common designs.
Munafò, M.R., Neill, J., 2016. Null is beautiful: On the importance of publishing null results. J. Psychopharmacol. (Oxf.) 30, 585–585. https://doi.org/10.1177/0269881116638813
Nickerson, R.S., 1998. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. Rev. Gen. Psychol. 2, 175–220. https://doi.org/10.1037/1089-2680.2.2.175
Nissen, S.B., Magidson, T., Gross, K., Bergstrom, C.T., 2016. Publication bias and the canonization of false facts. eLife 5, e21451. https://doi.org/10.7554/eLife.21451
Nolan, H., Whelan, R., Reilly, R.B., 2010. FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. J. Neurosci. Methods 192, 152–162. https://doi.org/10/fdmpw9
Nørgaard, M., Ganz, M., Svarer, C., Frokjaer, V.G., Greve, D.N., Strother, S.C., Knudsen, G.M., 2019. Different preprocessing strategies lead to different conclusions: A [11C]DASB-PET reproducibility study. J. Cereb. Blood Flow Metab. Off. J. Int. Soc. Cereb. Blood Flow Metab. 271678X19880450. https://doi.org/10.1177/0271678X19880450
Nosek, B.A., Bar-Anan, Y., 2012. Scientific Utopia: I. Opening Scientific Communication. Psychol. Inq. 23, 217–243. https://doi.org/10.1080/1047840X.2012.692215
Nosek, B.A., Beck, E.D., Campbell, L., Flake, J.K., Hardwicke, T.E., Mellor, D.T., van ’t Veer, A.E., Vazire, S., 2019. Preregistration Is Hard, And Worthwhile. Trends Cogn. Sci. https://doi.org/10.1016/j.tics.2019.07.009
Nosek, B.A., Ebersole, C.R., DeHaven, A.C., Mellor, D.T., 2018. The preregistration revolution. Proc. Natl. Acad. Sci. U. S. A. 115, 2600–2606. https://doi.org/10.1073/pnas.1708274114
Nosek, B.A., Spies, J.R., Motyl, M., 2012. Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability. Perspect. Psychol. Sci. 7, 615–631. https://doi.org/10.1177/1745691612459058
Oberauer, K., Lewandowsky, S., 2019. Addressing the theory crisis in psychology. Psychon. Bull. Rev. 26, 1596–1618. https://doi.org/10.3758/s13423-019-01645-2
Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Comput. Intell. Neurosci. 2011, 1–9. https://doi.org/10.1155/2011/156869
Open Science Collaboration, 2015. Estimating the reproducibility of psychological science. Science 349. https://doi.org/10.1126/science.aac4716
Palmer, C., Pe’er, I., 2017. Statistical correction of the Winner’s Curse explains replication variability in quantitative trait genome-wide association studies. PLOS Genet. 13, e1006916. https://doi.org/10.1371/journal.pgen.1006916
Papageorgiou, S.N., Xavier, G.M., Cobourne, M.T., Eliades, T., 2018. Registered trials report less beneficial treatment effects than unregistered ones: a meta-epidemiological study in orthodontics. J. Clin. Epidemiol. 100, 44–52. https://doi.org/10/gdzqnb
Pernet, C.R., Garrido, M., Gramfort, A., Maurits, N., Michel, C., Pang, E., Salmelin, R., Schoffelen, J.M., Valdes-Sosa, P.A., Puce, A., 2018. Best Practices in Data Analysis and Sharing in Neuroimaging using MEEG (preprint). Open Science Framework. https://doi.org/10.31219/osf.io/a8dhx
Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. Electroencephalogr. Clin. Neurophysiol. 72, 184–187. https://doi.org/10.1016/0013-4694(89)90180-6

Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafò, M.R., Nichols, T.E., Poline, J.-B., Vul, E., Yarkoni, T., 2017. Scanning the horizon: towards transparent and reproducible neuroimaging research. Nat. Rev. Neurosci. 18, 115–126. https://doi.org/10.1038/nrn.2016.167

Positively Negative: A New PLOS ONE Collection focusing on Negative, Null and Inconclusive Results, 2015. . PLOS Collect. URL https://collectionsblog.plos.org/positively-negative-new-plos-one-collection-focusing-negative-null-inconclusive-results-everyone/ (accessed 6.23.20).

Powell, K., 2018. How to sail smoothly from academia to industry. Nature 555, 549–551. https://doi.org/10/gc5tq3

Pyszczynski, T., Greenberg, J., 1987. Toward an Integration of Cognitive and Motivational Perspectives on Social Inference: A Biased Hypothesis-Testing Model, in: Advances in Experimental Social Psychology. Elsevier, pp. 297–340. https://doi.org/10.1016/S0065-2601(08)60417-7

Regan, D., 1977. Steady-state evoked potentials. J. Opt. Soc. Am. 67, 1475–1489. https://doi.org/10.1364/josa.67.001475

Rewarding negative results keeps science on track, 2017. . Nature 551, 414–414. https://doi.org/10.1038/d41586-017-07325-2

Robbins, K.A., Touryan, J., Mullen, T., Kothe, C., Bigdely-Shamlo, N., 2020. How Sensitive Are EEG Results to Preprocessing Methods: A Benchmarking Study. IEEE Trans. Neural Syst. Rehabil. Eng. Publ. IEEE Eng. Med. Biol. Soc. 28, 1081–1090. https://doi.org/10.1109/TNSRE.2020.2980223

Rosenthal, R., 1979. The file drawer problem and tolerance for null results. Psychol. Bull. 86, 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rouder, J.N., 2014. Optional stopping: No problem for Bayesians. Psychon. Bull. Rev. 21, 301–308. https://doi.org/10/f5xzcp

Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., Iverson, G., 2009. Bayesian t tests for accepting and rejecting the null hypothesis. Psychon. Bull. Rev. 16, 225–237. https://doi.org/10/b3hsdp

Sandre, A., Banica, I., Riesel, A., Flake, J., Klawohn, J., Weinberg, A., 2020. Comparing the effects of different methodological decisions on the error-related negativity and its association with behaviour and gender. Int. J. Psychophysiol. 156, 18–39. https://doi.org/10.1016/j.ijpsycho.2020.06.016

Scargle, J., 2000. Publication bias: the “File-Drawer” problem in scientific inference. J. Sci. Explor. 14, 91–106.

Scheel, A.M., Schijen, M., Lakens, D., 2020. An excess of positive results: Comparing the standard Psychology literature with Registered Reports (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/p6e9c

Schettino, A., Porcu, E., Gundlach, C., Keitel, C., Müller, M.M., 2020. Rapid processing of neutral and angry expressions within ongoing facial stimulus streams: Is it all about isolated facial features? PLOS ONE 15, e0231982. https://doi.org/10/gg89qr

Schindler, S., Schettino, A., Pourtois, G., 2018. Electrophysiological correlates of the interplay between low-level visual features and emotional content during word reading. Sci. Rep. 8, 12228. https://doi.org/10.1038/s41598-018-30701-5

Schönbrodt, F.D., Wagenmakers, E.-J., Zehetleitner, M., Perugini, M., 2017. Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. Psychol. Methods 22, 322–339. https://doi.org/10/gbkndx

Shepard, D., 1968. A two-dimensional interpolation function for irregularly-spaced data, in:

Proceedings of the 1968 23rd ACM National Conference On -. Presented at the the 1968 23rd ACM national conference, ACM Press, Not Known, pp. 517–524. https://doi.org/10.1145/800186.810616

Simmons, J.P., Nelson, L.D., Simonsohn, U., 2012. A 21 Word Solution. SSRN Electron. J. https://doi.org/10.2139/ssrn.2160588

Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychol. Sci. 22, 1359–1366. https://doi.org/10.1177/0956797611417632

Simonsohn, U., 2013. Just Post It: The Lesson From Two Cases of Fabricated Data Detected by Statistics Alone. Psychol. Sci. 24, 1875–1888. https://doi.org/10.1177/0956797613480366

Steegen, S., Tuerlinckx, F., Gelman, A., Vanpaemel, W., 2016. Increasing Transparency Through a Multiverse Analysis. Perspect. Psychol. Sci. 11, 702–712. https://doi.org/10.1177/1745691616658637

Swaen, G.G.M.H., Teggeler, O., van Amelsvoort, L.G.P.M., 2001. False positive outcomes and design characteristics in occupational cancer epidemiology studies. Int. J. Epidemiol. 30, 948–954. https://doi.org/10/dxb762

Szollosi, A., Donkin, C., 2019a. Arrested theory development: The misguided distinction between exploratory and confirmatory research (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/suzej

Szollosi, A., Donkin, C., 2019b. Neglected Sources of Flexibility in Psychological Theories: from Replicability to Good Explanations. Comput. Brain Behav. 2, 190–192. https://doi.org/10.1007/s42113-019-00045-y

Szollosi, A., Kellen, D., Navarro, D.J., Shiffrin, R., van Rooij, I., Van Zandt, T., Donkin, C., 2020. Is Preregistration Worthwhile? Trends Cogn. Sci. 24, 94–95. https://doi.org/10.1016/j.tics.2019.11.009

Szucs, D., Ioannidis, J.P.A., 2017. Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. PLOS Biol. 15, e2000797. https://doi.org/10.1371/journal.pbio.2000797

Topol, E.J., 2004. Failing the Public Health - Rofecoxib, Merck, and the FDA. N. Engl. J. Med. 351, 1707–1709. https://doi.org/10.1056/NEJMp048286

Toth, A.A., Banks, G.C., Mellor, D., O'Boyle, E.H., Dickson, A., Davis, D.J., DeHaven, A., Bochantin, J., Borns, J., 2020. Study Preregistration: An Evaluation of a Method for Transparent Reporting. J. Bus. Psychol. https://doi.org/10/ghcrqm

Trinquart, L., Dunn, A.G., Bourgeois, F.T., 2018. Registration of published randomized trials: a systematic review and meta-analysis. BMC Med. 16, 173. https://doi.org/10.1186/s12916-018-1168-6

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. Biol. Psychol. 72, 333–343. https://doi.org/10.1016/j.biopsycho.2005.11.012

Van den Akker, O., Weston, S.J., Campbell, L., Chopik, W.J., Damian, R.I., Davis-Kean, P., Hall, A.N., Kosie, J.E., Kruse, E.T., Olsen, J., Ritchie, S.J., Valentine, K.D., van 't Veer, A.E., Bakker, M., 2019. Preregistration of secondary data analysis: A template and tutorial (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/hvfmr

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N.J., Gronau, Q.F., Haaf, J.M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A.R.K.N., Sarafoglou, A., Stefan, A., Voelkel, J.G.,

Wagenmakers, E.-J., 2020. The JASP guidelines for conducting and reporting a Bayesian analysis. Psychon. Bull. Rev. https://doi.org/10/ghfv5h

van ’t Veer, A.E., Giner-Sorolla, R., 2016. Pre-registration in social psychology - A discussion and suggested template. J. Exp. Soc. Psychol., Special Issue: Confirmatory 67, 2–12. https://doi.org/10.1016/j.jesp.2016.03.004

VanRullen, R., 2011. Four Common Conceptual Fallacies in Mapping the Time Course of Recognition. Front. Psychol. 2. https://doi.org/10.3389/fpsyg.2011.00365

Vedula, S.S., Goldman, P.S., Rona, I.J., Greene, T.M., Dickersin, K., 2012. Implementation of a publication strategy in the context of reporting biases. A case study based on new documents from Neurontin® litigation. Trials 13, 136. https://doi.org/10.1186/1745-6215-13-136

Veldkamp, C.L.S., 2017. Doctoral thesis: The human fallibility of scientists - dealing with error and bias in academic research (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/g8cjq

Vul, E., Harris, C., Winkielman, P., Pashler, H., 2009. Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition. Perspect. Psychol. Sci. J. Assoc. Psychol. Sci. 4, 274–290. https://doi.org/10.1111/j.1745-6924.2009.01125.x

Wagenmakers, E.-J., 2007. A practical solution to the pervasive problems of p values. Psychon. Bull. Rev. 14, 779–804. https://doi.org/10/dgws9q

Wagenmakers, E.-J., Dutilh, G., 2016. Seven Selfish Reasons for Preregistration. APS Obs. 29.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H.L.J., 2011. Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). J. Pers. Soc. Psychol. 100, 426–432. https://doi.org/10.1037/a0022790

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H.L.J., Kievit, R.A., 2012. An Agenda for Purely Confirmatory Research. Perspect. Psychol. Sci. 7, 632–638. https://doi.org/10.1177/1745691612463078

Wasserstein, R.L., Lazar, N.A., 2016. The ASA Statement on p-Values: Context, Process, and Purpose. Am. Stat. 70, 129–133. https://doi.org/10.1080/00031305.2016.1154108

Wicherts, J.M., Veldkamp, C.L.S., Augusteijn, H.E.M., Bakker, M., van Aert, R.C.M., van Assen, M.A.L.M., 2016. Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. Front. Psychol. 7, 1832. https://doi.org/10.3389/fpsyg.2016.01832

Widmann, A., 2006. Firfilt EEGLAB Plugin. University of Leipzig, Leipzig.

Widmann, A., Schröger, E., 2012. Filter Effects and Filter Artifacts in the Analysis of Electrophysiological Data. Front. Psychol. 3. https://doi.org/10.3389/fpsyg.2012.00233

Widmann, A., Schröger, E., Maess, B., 2015. Digital filter design for electrophysiological data – a practical approach. J. Neurosci. Methods, Cutting-edge EEG Methods 250, 34–46. https://doi.org/10.1016/j.jneumeth.2014.08.002

Wilcox, R.R., 2016. Introduction to Robust Estimation and Hypothesis Testing, 4 edition. ed. Academic Press, Waltham, MA.

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., Teal, T.K., 2017. Good enough practices in scientific computing. PLOS Comput. Biol. 13, e1005510. https://doi.org/10.1371/journal.pcbi.1005510

Winkler, I., Haufe, S., Tangermann, M., 2011. Automatic Classification of Artifactual ICA-Components for Artifact Removal in EEG Signals. Behav. Brain Funct. 7, 30.

https://doi.org/10.1186/1744-9081-7-30
Wiseman, R., Watt, C., Kornbrot, D., 2019. Registered reports: an early example and analysis. PeerJ 7, e6232. https://doi.org/10.7717/peerj.6232
Xiao, R., Boehnke, M., 2009. Quantifying and correcting for the winner's curse in genetic association studies. Genet. Epidemiol. 33, 453–462. https://doi.org/10.1002/gepi.20398
Yamada, Y., 2018. How to Crack Pre-registration: Toward Transparent and Open Science. Front. Psychol. 9. https://doi.org/10/ggb8dx
Yarkoni, T., 2019. The Generalizability Crisis (preprint). PsyArXiv. https://doi.org/10.31234/osf.io/jqw35
Yarkoni, T., 2009. Big Correlations in Little Studies: Inflated fMRI Correlations Reflect Low Statistical Power - Commentary on Vul et al. (2009). Perspect. Psychol. Sci.