# Preregistration: Definition, advantages, disadvantages, and how it can help against questionable research practices

Angelos-Miltiadis Krypotos[1,2], Gaetan Mertens[3], Irene Klugkist[4], & Iris M. Engelhard[1]

[1] Department of Clinical Psychology, Utrecht University

[2] Group of Health Psychology, KU Leuven

[3] Medical and Clinical Psychology, Tilburg University

[4] Methodology and Statistics, Utrecht University

**Preprint version of the chapter available at the book Avoiding Questionable Research Practices in Applied Psychology ( https://link.springer.com/book/10.1007/978-3-031-04968-2). Please read the final version on the Chapter as there maybe differences between versions.**

AMK is supported by a senior post-doctoral grant from FWO (Reg. # 12X5320N) and a replication grant from the Netherlands Organization for Scientific Research (NWO) (Reg. # 401.18.056). Iris M. Engelhard is supported with a Vici innovational research grant by NWO (grant number: 453-15-005).

PREPRINT

Correspondence concerning this article should be addressed to Angelos-Miltiadis Krypotos, Martinus J. Langeveldgebouw, Heidelberglaan 1, Room H.131, 3584 CS Utrecht, the Netherlands. E-mail: amkrypotos@gmail.com.

PREPRINT

Abstract

Questionable Research Practices (QRPs), such as p-hacking (i.e., the inappropriate manipulation of data analysis to find statistical significance) and post-hoc hypothesizing, are threats to replicability of research findings. One key solution to the problem of QRPs is pre-registration. This refers to time-stamped documentation that describes the methodology and statistical analyses of a study before the data are collected or inspected. As such, readers of the study's report can evaluate whether the described research is in line with the planned methods and analyses or there are deviations from these (e.g., analyses performed so that the research hypotheses is confirmed). Here, we aim to describe what preregistration entails and why it is useful for psychology research. In this vein, we present the key elements of a sufficient preregistration file, its advantages, as well as its disadvantages, and why preregistration is a key, yet partially insufficient, solution against QRPs. By the end of this chapter, we hope that readers are convinced that there is little reason not to preregister their research.

PREPRINT

# Preregistration: Definition, advantages, disadvantages, and how it can help against questionable research practices

Credible psychological science implies that research is reproducible or replicable. Concerns about whether the psychology literature is reliable have been raised a long time ago (Babbage, 1830; Rosenthal, 1979; Stroebe, Postmes, & Spears, 2012). However, this past decade, a crisis in the confidence of psychology, as well as other scientific fields has risen (Camerer et al., 2016; National Academies of Sciences & Medicine, 2019; Pashler & Wagenmakers, 2012). This was primarily due to the publication of studies showing poor replicability (the repetition of a study's findings with new data) of many important psychological findings (Open Science Collaboration, 2015; Ritchie, 2020), as well as poor reproducibility (finding of identical results when performing the original analyses on the same data) of research (Hardwicke et al., 2020, 2019).

One of the proposed reasons for the low replicability and reproducibility in psychology is Questionable Research Practices (QRPs). QRPs include the formation of a research hypothesis after the results are known (HARKing; Kerr, 1998), the flexible use of data analyses to obtain evidence for a hypothesis, even when it is not supported by the data (Simmons, Nelson, & Simonsohn, 2011), and the collection of data until the null hypothesis is rejected in Null-Hypothesis Significance Testing (NHST; Strube, 2006). The reported high prevalence of QRPs in psychology (John, Loewenstein, & Prelec, 2012), demands immediate changes in our research practices and the establishment of ways that prevent researchers from using these.

Diverse methods for eliminating QRPs have been proposed. The first method is to change the incentive structures in science (Bruton, Medlin, Brown, & Sacco, 2020; Chambers, Dienes,

PREPRINT

McIntosh, Rotshtein, & Willmes, 2015). In particular, academic success is commonly evaluated by the number of articles scientists have published in journals with high impact factors. Given that the report of significant results increases the chances that a paper will be published (Fanelli, 2012; Rosenthal, 1979), QRPs bias the results towards this direction (Fanelli, 2010). These QRPs include the deviation of data collection procedures so that the results would support the predictions made by the authors, the removal of data without a justifiable reason, or even data fabrication. Different proposals have been made to solve this problem, such as a training in ethics in science (Bruton et al., 2020) or emphasizing science quality as indicator of academic success.

A second call towards tackling QRPs is the open sharing of data and material as well as the replication of past findings. Reproduction, however, is often challenging given that researchers typically do not share their data and material broadly (Alsheikh-Ali, Qureshi, Al-Mallah, & Ioannidis, 2011; Hardwicke et al., 2020, 2019; Vanpaemel, Vermorgen, Deriemaecker, & Storms, 2015; Vines et al., 2014). Replication of past findings had also been done sparingly, given that replication studies are traditionally harder to publish compared to original findings. To date, however, more and more journals (e.g., *Journal of Experimental Psychology: General*) and funding agencies (e.g., the Netherlands Organisation for Scientific Research) call for such studies, giving hope that this will be a way to reduce QRPs (Zwaan, Etz, Lucas, & Donnellan, 2018).

A third way to battle QRPs is the *preregistration* of a study before data analysis. The preregistration of studies is not new: the first registries were introduced in the 1960s (see Wiseman, Watt, & Kornbrot, 2019 for a full historical review). To date, preregistration is routinely done in some scientific fields (e.g., for clinical trials; see clinicaltrials.gov in the United States of America and eudract.ema.europa.eu in Europe). However, there is a call for the

PREPRINT

extensive preregistration in psychology for all experimental studies, meta-analyses, and literature reviews. Preregistration is increasingly used (Lindsay, Simons, & Lilienfeld, 2016; Nosek & Lindsay, 2018; J. Simmons et al., 2021b). It is promoted by journals, for instance, by providing a badge to a published article if the reported study was preregistered (Kidwell et al., 2016), or by not allowing a paper to be published unless the authors preregistered the research or explain why they did not. Also, more and more journals are requiring (e.g., *The Journal of Politics*) or at least encourage (e.g., *PAIN*) the preregistration of experimental studies. Similar strategies are also encouraged for researchers and graduate students (e.g., in the Behavioural Science Institute of Radboud University) and the same goes for some grant agencies (e.g., ZonMw in the Netherlands).

Preregistration is the topic of this chapter. Specifically, we aim to explain what preregistration is, why it is useful, what its shortcomings are, as well as why it can provide a shield against some QRPs.

The structure of this chapter is as follows: We first describe what preregistration is, and the key distinctions in the type of preregistrations. Then, we provide key sections that are typically included in a preregistration document. Furthermore, we discuss the advantages and challenges of preregistration and end by providing alternatives to preregistration. At the end of this chapter, we hope that readers will be convinced that it is imperative to preregister their research.

PREPRINT

## 1 What is Preregistration?

Preregistration consists of a collection of time stamped documents that typically describe a study's research questions, hypotheses, methodology, and statistical analyses. Although a study

should be preregistered before the beginning of data collection, this is not always possible (e.g., see below about preregistration of pre-existing data), As such, as a general rule a study should be preregistered at least *before* the research data are inspected (Nosek et al., 2018).

Multiple preregistration templates have been introduced (e.g., Johnson & Cook, 2019; Krypotos, Klugkist, Mertens, & Engelhard, 2019; Mertens & Krypotos, 2019), with different criteria to be fulfilled depending on the type of study (e.g., meta-analysis, laboratory studies, single-case designs), or when the study was preregistered (e.g., prior to data collection or to the data analysis). As such, researchers should first define clearly under which category their study falls.

A widely accepted distinction in preregistration is between studies in which original data are collected (e.g., laboratory research or a randomized controlled trial) and studies that use pre-existing data (e.g., re-analysis of an available data set). In the former case, researchers should define their research questions, hypotheses (if any), methods, and analyses. In the latter case, the preregistration of methods is more limited as the data have been already collected. In case of pre-existing data, researchers should acknowledge that they have already, at least partial, knowledge of the data set to be analyzed, which could influence their analytic choices. Another commonly used distinction is whether a study aims to confirm a hypothesis (i.e., *confirmatory research*) or whether its goal is to explore different data patterns, without having an a-priori hypothesis or a computational model to confirm (i.e., *exploratory research*) (De Groot, 2014; Dirnagl, 2020). Notably, the distinction in the different types of studies is not qualitative, as the different types of studies serve different purposes. For example, exploratory studies may enable the development of novel computational models, whereas confirmatory studies are needed to provide supportive evidence or gain more confidence in, the structure of a particular pre-specified model.

PREPRINT

Despite different preregistration templates for different studies, there are many commonalities in preregistration templates. Below, we will present the common aspects of preregistration, and we will describe which deviations are needed for different types of studies.

## 2 Advantages of Preregistration.

There are plenty of reasons to preregister a study. Here we provide some of the key advantages of preregistration, before moving to a series of challenges in the next section.

First, preregistration allows researchers to take full credit for making an accurate prediction. Think, for example, of someone pulling random numbers from a bag. She picks one number, sees it, and claims "It is number 4 as I had predicted". Without her having mentioned her prediction publicly in advance, her claim is not strong enough. Preregistration allows scientists to take full credit for the accuracy of their predictions by providing clear evidence that these were made in advance.

Second, in line with open science (Allen & Mehler, 2019), preregistration is away to show that you are conducting transparent research, with results that are not based on post-hoc reasoning and analyses (see QRPs above) but with concrete predictions made in advance. Increasingly, science funders and journals require researchers to demonstrate that their research practices are in line with open science principles. Preregistration is one way to achieve this.

Third, from a philosophy of science standpoint, preregistration allows researchers to transparently evaluate the *severity of their tests* (Mayo, 2018). Dating back from the time of Sir Karl Popper (e.g., Popper, 2005), a test is argued to be severe when it is strong enough to falsify a theory. In this line of reasoning, a preregistration allows others to evaluate whether a performed

PREPRINT

test was capable of falsifying a tested theory (Hitzig & Stegenga, 2020; Lakens, 2020; O'Donohue, 2021; Vanpaemel, 2019).

From a practical point of view, preregistration allows researchers to wrap up a project faster compared to when they have to decide all analytic options after data collection. Also, in case of results that do not confirm someone's hypothesis, there is the temptation to abandon a project altogether. By preregistrating the study, researchers already have the basic material for writing their methods/analysis section, and a specific plan for carrying out all the analyses. Lastly, preregistration may also help researchers in protecting themselves against unwarranted requests for additional data analyses by reviewers, which can delay the publication of their results.

Despite the advantages of preregistrating a study, there are also arguments against preregistration of (some) studies. We turn to these below.

## 3 Disadvantages of Preregistration (and how to counter them).

No scientific practice is without its shortcoming, and as such preregistration is not without its shortcoming (e.g., Rubin, 2020). We will discuss 8 disadvantages below and will show that they are less important than the relative advantages.

First, it is not uncommon that by the end of data collection, researchers have thought of a different, and better, way to analyze their data than the way they mentioned in the preregistration. Also, a new statistical method may have been introduced between the preregistration and the end of data collection. As mentioned above, in these cases the authors may update their

PREPRINT

preregistration by providing the reasons for applying a new analysis and mentioning the reasons why such analysis is superior to the preregistered ones.

The second disadvantage relates to the limits of preregistration although it makes research design and analysis plans transparent but not necessarily correct or relevant. To illustrate, even when a researcher preregisters that she is going to perform a paired-samples $t$-test for group comparisons, this does not mean that such an analysis is correct. In this specific example, a paired-samples $t$-test would be not be the right option as one of its basic assumptions is that each pair of observations come from the same participant/group, making the between group comparisons impossible. A solution to this disadvantage could be given in the form of Registered Reports (Chambers, 2013). This type of article includes the evaluation of the study's introduction and methods *prior* to the beginning of data collection. The reviewers can evaluate the soundness of the methodology, the statistical analyses, as well as the relevance of the study for the specific journal. After the paper has been accepted as Registered Report, the authors can collect the data and resubmit the article. The Registered Report format also protects researchers from reviewers' critiques after the results are known (Wagenmakers & Dutilh, 2016). Registered Reports are currently adopted by almost 300 journals (see https://www.cos.io/initiatives/registered-reports for the full list of journals).

Third, preregistration calls for a change in the workflow of doing research, which could be particularly difficult especially for seasoned researchers. In order to ensure that researchers use preregistration in their work, many relevant user-friendly programs have been introduced (e.g., see Krypotos et al., 2019).

PREPRINT

Fourth, there is an ongoing discussion as to whether preregistration is worthwhile in the first place (Nosek et al., 2019; Szollosi et al., 2019). This point relates to the idea that preregistration improves the diagnostic value of the statistical tests (see also severity tests above). For example, preregistration is argued to enable an accurate familywise error rate (i.e., the probability of making at least one false discovery when running multiple statistical tests) and to force people to think deeply about their theories (Nosek et al., 2019). Still, these ideas have been challenged (Olken, 2015; Szollosi et al., 2019) and a call for better theories have been made instead of the ubiquitous adoption of preregistration.

Fifth, it has been argued that preregistration cannot really limit QRPs, as it may give them a different form, such as preregistration after the study has been completed (Yamada, 2018). Related to that, there has been a misuse of the badges awarded to some studies, with some articles reporting multiple studies and gaining badges for only preregistrating part of the studies (Claesen, Gomes, Tuerlinckx, & Vanpaemel, 2019). Still, such disadvantages do not relate to the limitations of preregistration as a tool, but to its misuse by researchers.

Sixth, exploratory research is sometimes considered to be less strong compared to confirmatory research, so a concern could be that ‘safer’ research will be promoted that is focused on the confirmation of largest effects, and that exploratory research is put in the second place (Pham & Oh, 2020). This argument relates to the faulty misconception that exploratory research is a second-tier research, although it is equally important as confirmatory research (Scheel, Tiokhin, Isager, & Lakens, 2020). Preregistration just helps researchers to better separate these two types of research, but it does not value one as better than the other (Simmons et al., 2021b; J. Simmons et al., 2021a).

PREPRINT

Seventh, the preregistration does not just concern the authors but also the reviewers and editors. It is important that the reviewers and editors carefully confirm that the preregistration plan has been followed and if not, that the deviations are reported in the manuscript. Although this may seem like a lot of work (Pham & Oh, 2020), in principle it will result in less work as the reviewers or editors do not have to question whether the results were p-hacked as preregistration plan has been shown (Wagenmakers & Dutilh, 2016).

Lastly, it is tempting not to follow the preregistration plan, especially when the research results are against the hypotheses of the study. Sadly, there is evidence that often the published results differ from the preregistration plan (Claesen et al., 2019). However, this is not a disadvantage of preregistration per se but of the current incentive system in science.

## 4 What to Include in a Study's Preregistration

### 4.0.1 Research hypotheses..

Confirmatory research is conducted for proving or falsifying a hypothesis (see O'Donohue, 2021 for a discussion of this issue). As such, specific hypotheses should be determined explicitly in advance in the preregistration. General research hypotheses may leave too much room for flexible data analyses. In contrast, exploratory research does not require explicit or specific hypotheses.

### 4.0.2 Methodology..

Following the research hypotheses, the methodology for testing the hypotheses should be described. Although the methodology is more extensive when original data are collected (see below), even studies with preexisting data should include a methodology section, including how the data were acquired, or in case of a meta-analysis, how these will be retrieved and extracted from the literature. Notably, in case the data have already

PREPRINT

been published, a link to the previous research should be included, and prior information about the data should be disclosed that could influence the analytic decisions (see below).

The methodology of a study includes the definition of (if applicable): stimuli that will be used, questionnaires and answering scales, procedures, blinding of the experimenters, and randomization. In line with open-science practices, it would be desirable if all relevant materials are uploaded in a repository, so other researchers have access to all original materials in case they want to replicate the study.

**4.0.3 Sample..** The characteristic of the (intended) sample should be described in the preregistration document. Although sex and age descriptions are standard in psychology research, other sample characteristics that are relevant to the research questions should be included as well. For example, a study regarding anxiety disorders, may also include anxiety levels of the sample. Notably, characteristics of the sample can influence the generalizability of results to other samples. This is particularly important, because current samples in psychology mostly include Western, Educated, Industrial, Rich, and Democratic (WEIRD) samples [Henrich, Heine, and Norenzayan (2010); Muthukrishna et al. (2020); see chapter XX this volume], that often limits the generalizability of the findings to other populations.

An important decision that has to be made before the beginning of the study is the size of the sample. There are different ways to justify the sample size of a study. For example, the researcher could run a power analysis (Cohen, 1992) based on the effect sizes being previously reported in the literature or by defining the effect size that is minimally interesting for a study (i.e., *the minimal statistically detectable effect*) (Albers & Lakens, 2018; Lakens, 2020). This is the minimum effect that if present, would be statistically significant given the sample size of the

PREPRINT

study and the chosen $\alpha$ level (Cook et al., 2014). Importantly, the size of the sample has an important influence on the direction of the results, especially when the analyses are run within a Null-Hypothesis Significance Testing (NHST) Framework. Within a NHST framework, $p$-values (i.e., the probability of observing the current or more extreme data given that the null-hypothesis is true; Wagenmakers, 2007) will almost always turn out to be statistically significant given when that enough sample data are collected. This is also the case when, even if the tested effects comes from population correctly as described by the null-hypothesis.

As an alternative to defining the sample size in advance, it is also possible to use adaptive procedures. In these procedures, data collection is completed when adequate evidence has been accumulated for or against a hypothesis or it is completed based on other objective criteria, such as the time the lab is available. For example, an investigator could use sequential analyses, where the $\alpha$ level is divided by the times a test is planned to be performed (Lakens, 2014), or use a Bayesian data planning procedure (Schönbrodt & Wagenmakers, 2018). Our goal here is to make it explicit that no matter which stopping rule is used in the study, this should be mentioned clearly in advance in a preregistration document.

In case of pre-existing data sets, the sample characteristics that need to be reported depend on the research question. To illustrate, in case of a genome-wide association study, which is conducted to test whether specific genes predict the development of psychopathology, the preregistration should include only the characteristics of the subset of the sample. When a new model will be tested, there may be a distinction between the data that are used for tuning the model parameters and validating the model (also referred as the *training* and *validating* data set in machine learning; Dwyer, Falkai, and Koutsouleris (2018)). If this is the case, then the preregistration should mention how the two separate data sets will be determined.

PREPRINT

**4.0.4 Data preprocessing..** Before the data analyses, scientists often transform their data, reject outliers, and aggregate values to sum or mean scores. For example, in case of reaction time (RT) analyses, extreme values are typically removed and the distribution of RTs is log transformed (Heathcote, Popiel, & Mewhort, 1991). It is important that all data transformations are also included in the preregistration given that different transformation may change the direction of the results. Whether each choice is defendable or not is up to the researcher and the scientific community. Still, the modification of the data may determine the direction of results, and non-specification of data transformation/reduction processes leaves room for QRPs. In some research fields, however, exact predefining the data reduction/transformation procedures is almost impossible, given that such procedures are often dictated by the data per se (e.g., normalize distribution of data only if they show that they are distributed normally). In such cases, researchers are advised to list the sensitivity analyses they will perform to ensure that the direction of the findings are not the result of the data reduction procedure.

**4.0.5 Statistical analysis..** Statistical analyses follow from the theoretical background of the study and the research questions. In cases of concrete formal theories, the statistical analyses follow such models and researchers have reduced flexibility in choosing which analysis they should perform (Rooij & Blokpoel, 2020). However, such formal models are rare in psychology, and usually generic statistical models for drawing inferences are selected, such as regression, *t*-tests, analysis of variance (ANOVA). Nonetheless, and due to the absence of a formal model (Rooij & Baggio, 2021), the same research question can be answered with different analyses. To illustrate, during a fear conditioning task, in which initially neutral stimuli are paired with unpleasant stimuli across multiple trials, someone could run a repeated measures

PREPRINT

ANOVA or a multilevel model. Given that different analyses can yield different results, flexibility in such statistical analyses may inflate false positive rates (Simmons et al., 2011).

To convince readers that the analyses were free from biases stemming from data inspection, a clear description of the planned statistical analyses should be included in the preregistration document. This description includes the inferential framework (e.g., NHST, Bayesian analyses) and the statistical models that will be used, with a clear definition of the variables in the model. Arguably, many decisions cannot be taken without inspecting the data and some data reduction procedures cannot be predicted (e.g., a data pattern against expectations). There are at least two ways to solve problems with decisions that need to be made before data inspection. First, researchers can create a flow chart of how the data could inform the statistical decisions. For instance, someone could argue that when the assumption of a normality in the variables is violated, a Welch's *t*-test will be used instead of a Student's *t*-test. Accounting for each possible data pattern will be daunting, especially when complex statistical models are used. An alternative strategy would be to update the preregistration file and to argue why the newly proposed analyses are a better approach to the data analyses. Preregistration should be viewed as a plan, not a prison (DeHaven, 2017), that can be updated. Such updates should be shared timely and transparently with the rest of the community so they can be judged accordingly.

**4.0.6 Remaining sections..** Above, we described common characteristics of current preregistration templates (e.g., Crüwell & Evans, 2019; Kirtley, Lafit, Achterhof, Hiekkaranta, & Germeys, 2020; Krypotos et al., 2019; Mertens & Krypotos, 2019; Van den Akker et al., 2019; Veer & Giner-Sorolla, 2016). As mentioned previously, however, different types of studies need different preregistration elements, and more templates are being introduced

PREPRINT

depending on the field of study. We suggest that authors first inspect available templates and choose the template that best fits their study [see https://osf.io/zab38/ for an overview].

## 5 Where to preregister

The completed preregistration document should be submitted to an official repository. To date, most repositories are online. The type of repository that will be used also depends on the type of study. For example, clinical trials are most commonly registered in clinicaltrials.gov in the United States and eudract.ema.europa.eu for Europe. For experimental and modelling work within psychology, two databases are commonly used. The first one is *aspredicted* (aspredicted.org). It enables researchers to preregister a study by answering 9 simple questions relating to the study's research design and analyses. The second one is *osf* (osf.io) where researchers have the option to select templates in which they answer many more questions compared to aspredicted, and go much more into depth in their study. We urge authors to prefer including enough information in their study compared to vague specification, something that will leave less room for misinterpretations as well as flexible data analyses. A limitation of the aspredicted website is that although the preregistration is quite easy to complete, the website is *not* a formal registry given that preregistrations could be kept private forever. In contrast, on the osf, preregistrations are released online after a maximum of four years. Allowing researchers to keep their preregistration private could result in preregistrating multiple hypotheses and releasing only the ones that support the preregistration file that supports their study.

PREPRINT

## 6 Alternatives to Preregistration.

Preregistration is only one way to counter QRPs. In this section, we will suggest additional tools that could be used in combination with preregistration, although these are not integral parts of preregistration.

The first one is crowdsourcing analyses. This includes the sharing of a dataset with different groups that are allowed to analyze the data set in any desired way (Dutilh et al., 2019; Silberzahn et al., 2018). To illustrate, in Dutilh, Annis, et al. (2019), the first author shared the same set data with different groups of experts on diffusion models, a computational model used for decomposing reaction time performance into different model parameters (Ratcliff & McKoon, 2008). The different groups then had to fit any form of the diffusion model (i.e., a computational model used for decomposing reaction time performance into different model parameters Ratcliff & McKoon, 2008) and report back model parameter values. This allowed the different groups to use any version of the diffusion model they wanted, with no two groups selecting the same model, without knowing explicitly what was the research question. This approach can reduce the bias towards presenting results supporting an effect.

The second alternative is a multiverse analysis (Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016). The rationale of multiverse analyses is that in the absence of a concrete background, more than one type of analyses seems reasonable. Let us return to the reaction time example. Reaction time distributions are typically skewed. Typically, a summary value is used for a reaction time distribution (e.g., the mean). It could be argued that a given researcher prefers computing the median of the distribution, because it is less influenced by extreme values compares to the mean. Another approach would be to normalize the distribution and then

PREPRINT

compute the mean. In absence of a theory about the best option, both options are reasonable. In multiverse analyses, researchers need to conduct all reasonable analyses. Then, the distribution of results is plotted. Multiverse analyses can be specified on the level of data reduction procedure (e.g., different data transformation) but also on the level of the selected statistical models (e.g., multilevel analyses or Analyses of variance).

## 7 Concluding remarks.

To date, preregistration of a study constitutes one of the most important tools towards battling QRPs. As shown above, however, it does not provide absolute immunity towards them. Nevertheless, given the advantages presented above, there is little reason not to preregister a study. It is likely that in the next years, preregistration of a study will become the norm, rather than an exception, and it is possible that over the next decade, there will be hardly any experimental study in our field that is not preregistered. This norm, however, can have exceptions, and as such researchers can always simply argue as to why they did not preregister their study.

In order to achieve the goal that science is open, transparent, and replicable, we will need to move towards adopting better practices, such as the open sharing of data and materials. Ultimately, the goal of science is the collection of reliable information that is useful for science itself, and the whole society. QRPs do not serve that goal and should be maximally eliminated, such as by the adoption of study preregistration.

PREPRINT

## 8 References

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, *74*, 187–195.

Allen, C., & Mehler, D. M. (2019). Open science challenges, benefits and tips in early career and beyond. *PLoS Biology*, *17*, e3000246.

Alsheikh-Ali, A. A., Qureshi, W., Al-Mallah, M. H., & Ioannidis, J. P. (2011). Public availability of published research data in high-impact journals. *PloS ONE*, *6*, e24357.

Babbage, C. (1830). *Reflections on the decline of science in england: And on some of its causes, by charles babbage (1830).* (Vol. 1). B. Fellowes.

Bruton, S. V., Medlin, M., Brown, M., & Sacco, D. F. (2020). Personal motivations and systemic incentives: Scientists on questionable research practices. *Science and Engineering Ethics*, 1–17.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., … others. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*, 1433–1436.

Chambers, C. D. (2013). Registered reports: A new publishing initiative at cortex. *Cortex*, *49*, 609–610.

Chambers, C. D., Dienes, Z., McIntosh, R. D., Rotshtein, P., & Willmes, K. (2015). Registered reports: Realigning incentives in scientific publishing. *Cortex*, *66*, A1–A2.

PREPRINT

Claesen, A., Gomes, S. L. B. T., Tuerlinckx, F., & Vanpaemel, W. (2019). Preregistration: Comparing dream to reality.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*, 98–101.

Cook, J. A., Hislop, J., Adewuyi, T. E., Harrild, K., Altman, D. G., Ramsay, C. R., … others. (2014). Assessing methods to specify the target difference for a randomised controlled trial: DELTA (difference elicitation in trials) review. *Health Technology Assessment (Winchester, England)*, *18*.

Crüwell, S., & Evans, N. J. (2019). Preregistration in complex contexts: A preregistration template for the application of cognitive models. *Preprint Available at PsyArXiv*.

De Groot, A. (2014). The meaning of "significance" for different types of research [translated and annotated by Eric-Jan Wagenmakers, Denny Borsboom, Josine Verhagen, Rogier Kievit, Marjan Bakker, Angelique Cramer, Dora Matzke, Don Mellenbergh, and Han LJ van der Maas]. *Acta Psychologica*, *148*, 188–194.

DeHaven, A. (2017). Preregistration: A plan, not a prison. *Retrieved October*, *29*, 2019.

Dirnagl, U. (2020). Preregistration of exploratory research: Learning from the golden age of discovery. *PLoS Biology*, *18*, e3000690.

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P., … others. (2019). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. *Psychonomic Bulletin & Review*, *26*(4), 1051–1069.

PREPRINT

Dutilh, G., Sarafoglou, A., & Wagenmakers, E.-J. (2019). Flexible yet fair: Blinding analyses in experimental psychology. *Synthese*, 1–28.

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, *14*, 91–118.

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PloS One*, *5*, e10271.

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*, 891–904.

Hardwicke, T. E., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M. B., Peloquin, B., … Frank, M. C. (2020). Analytic reproducibility in articles receiving open data badges at psychological science: An observational study. *Royal Society Open Science*.

Hardwicke, T. E., Wallach, J. D., Kidwell, M. C., Bendixen, T., Crüwell, S., & Ioannidis, J. P. (2019). An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014–2017). *Royal Society Open Science*, *7*, 190806.

Heathcote, A., Popiel, S. J., & Mewhort, D. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, *109*, 340.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*, 29–29.

Hitzig, Z., & Stegenga, J. (2020). The problem of new evidence: P-hacking and pre-analysis plans. *Diametros*, *17*, 10–33.

PREPRINT

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524–532.

Johnson, A. H., & Cook, B. G. (2019). Preregistration in single-case design research. *Exceptional Children*, *86*, 95–112.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*, 196–217.

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., … others. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology*, *14*, e1002456.

Kirtley, O., Lafit, G., Achterhof, R., Hiekkaranta, A., & Germeys, I. (2020). Making the black box transparent: A template and tutorial for (pre-) registration of studies using experience sampling methods (ESM). *Advances in Methods and Practices in Psychological Science*.

Krypotos, A.-M., Klugkist, I., Mertens, G., & Engelhard, I. M. (2019). A step-by-step guide on preregistration and effective data sharing for psychopathology research. *Journal of Abnormal Psychology*, *128*, 517–527.

Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology*, *44*, 701–710.

Lakens, D. (2020). Sample size justification. *Preprint Available at PsyArXiv*.

Lindsay, D. S., Simons, D. J., & Lilienfeld, S. O. (2016). Research preregistration 101. *APS Observer*, *29*.

PREPRINT

Mayo, D. G. (2018). *Statistical inference as severe testing*. Cambridge: Cambridge University Press.

Mertens, G., & Krypotos, A.-M. (2019). Preregistration of analyses of preexisting data. *Psychologica Belgica*, *59*, 338.

Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond western, educated, industrial, rich, and democratic (weird) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychological Science*, 0956797620916782.

National Academies of Sciences, Engineering, & Medicine. (2019). *Reproducibility and replicability in science*. National Academies Press.

Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., … Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences*, *23*, 815–818.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*, 2600–2606.

Nosek, B. A., & Lindsay, D. S. (2018). Preregistration becoming the norm in psychological science. *APS Observer*, *31*(3).

O'Donohue, W. (2021). Some Popperian notes regarding replication failures in psychology.

Olken, B. A. (2015). Promises and perils of pre-analysis plans. *Journal of Economic Perspectives*, *29*, 61–80.

PREPRINT

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*.

Pashler, H., & Wagenmakers, E. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*, 528–530.

Pham, M. T., & Oh, T. T. (2020). Preregistration is neither sufficient nor necessary for good science. *Journal of Consumer Psychology*.

Popper, K. (2005). *The logic of scientific discovery*. Routledge.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.

Ritchie, S. (2020). Science fictions: How fraud, bias, negligence, and hype undermine the search for truth. London: The Bodley Head.

Rooij, I. van, & Baggio, G. (2021). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science*.

Rooij, I. van, & Blokpoel, M. (2020). Formalizing verbal theories. *Social Psychology*, *51*, 285–298.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 638–641.

Rubin, M. (2020). Does preregistration improve the credibility of research findings? *The Quantitative Methods in Psychology*, *16*, 376–390.

PREPRINT

Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2020). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science*.

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*, 128–142.

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., … others. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, *1*, 337–356.

Simmons, J., Nelson, L., & Simonsohn, U. (2021a). Pre-registration is a game changer. But, like random assignment, it is neither necessary nor sufficient for credible science. *Journal of Consumer Psychology*.

Simmons, J., Nelson, L., & Simonsohn, U. (2021b). Pre-registration: Why and how. *Journal of Consumer Psychology*.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*, 702–712.

Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, *7*, 670–688.

PREPRINT

Strube, M. J. (2006). SNOOP: A program for demonstrating the consequences of premature and repeated null hypothesis testing. *Behavior Research Methods*, *38*, 24–27.

Szollosi, A., Kellen, D., Navarro, D. J., Shiffrin, R., Rooij, I. van, Van Zandt, T., & Donkin, C. (2019). Is preregistration worthwhile. *Trends in Cognitive Sciences*, *24*, 94–95.

Van den Akker, O., Weston, S. J., Campbell, L., Chopik, W. J., Damian, R. I., Davis-Kean, P., … others. (2019). Preregistration of secondary data analysis: A template and tutorial.

Vanpaemel, W. (2019). The really risky registered modeling report: Incentivizing strong tests and HONEST modeling in cognitive science. *Computational Brain & Behavior*, *2*, 218–222.

Vanpaemel, W., Vermorgen, M., Deriemaecker, L., & Storms, G. (2015). Are we wasting a good crisis? The availability of psychological research data after the storm. *Collabra: Psychology*, *1*.

Veer, A. E. van't, & Giner-Sorolla, R. (2016). Pre-registration in social psychology—a discussion and suggested template. *Journal of Experimental Social Psychology*, *67*, 2–12.

Vines, T. H., Albert, A. Y., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., … Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology*, *24*, 94–97.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p*-values. *Psychonomic Bulletin & Review*, *14*, 779–804.

Wagenmakers, E.-J., & Dutilh, G. (2016). Seven selfish reasons for preregistration. *APS Observer*, *29*.

PREPRINT

Wiseman, R., Watt, C., & Kornbrot, D. (2019). Registered reports: An early example and analysis. *PeerJ*, *7*, e6232.

Yamada, Y. (2018). How to crack pre-registration: Toward transparent and open science. *Frontiers in Psychology*, *9*, 1831.

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*.

PREPRINT