Preregistration of Studies with Existing Data

36

Gaëtan Mertens and Angelos-Miltiadis Krypotos

Abstract

Preregistration of research plans is becoming an increasingly popular and common tool to enhance the transparency of a study's methodology. In a preregistration, researchers document their research plans and register them to a public repository prior to conducting their research. In this chapter, we provide arguments for why preregistration can protect scientific findings against questionable research practices (QRPs), such as outcome swapping, selective reporting of conditions, unwarranted data exclusions, and post hoc changing of hypotheses. Furthermore, we place particular emphasis on preregistering research plans when using existing data, and we give an overview of preregistration templates and public repositories for different types of research designs. We conclude this chapter with highlighting some of the common criticisms of preregistration and our counterarguments and provide future reflections.

List of Abbreviations

HARKing	Hypothesizing after the results are known
NHST	*Null hypothesis significance testing*
QRPs	Questionable research practices

36.1 Introduction

Science relies for a large part on the collection and analysis of empirical data. Within the dominant hypothetico-deductive model of science, empirical data may be used to generate novel theories or test existing theories [1]. The underlying idea is that less accurate theories to predict and explain empirical data are gradually replaced with new theories that are more accurate or simpler than the old theories.

Until recently, researchers nearly always had to collect new empirical data to test their theories. However, due to rapid advances in the capacity to easily store and share data on online servers, datasets are now easily available to researchers across the world. While this facilitates the work of scientists, it also brings about new challenges to ensure accurate inferences based on data. To illustrate, can hypotheses still be validly tested on a dataset that has been collected for different

Gaëtan Mertens and Angelos-Miltiadis Krypotos are contributed equally

G. Mertens (✉)
Department of Medical and Clinical Psychology, Tilburg University, Tilburg, The Netherlands
e-mail: g.mertens@tilburguniversity.edu

A.-M. Krypotos
Leuven Research Group on Health Psychology, KU Leuven, Leuven, Belgium

Department of Clinical Psychology, Utrecht University, Utrecht, The Netherlands
e-mail: a.m.krypotos@uu.nl

J. Faintuch, S. Faintuch (eds.), *Integrity of Scientific Research*,
https://doi.org/10.1007/978-3-030-99680-2_36

purposes? And how can it be ensured that the hypothesis was really specified independently of the data? Alternatively, are existing data merely useful for exploratory research (i.e., finding patterns in the data by the reanalysis of data)? In this chapter, we aim to answer these questions regarding the reuse of existing data to test hypotheses. First, however, we outline how the validity of hypothesis testing can be threatened through the use of questionable research practices.

36.2 Threats to the Validity of Scientific Inferences: Questionable Research Practices (QRPs)

Typically, when scientists want to test a theory, they will propose a falsifiable hypothesis and test this with empirical data [2]. Because empirical data are nearly always influenced by random noise, statistical models are applied to the data to quantify the reliability of the observations. In addition, statistical inference frameworks are used to deduce the population distributions from the collected sample data. Such deductions, as well as the correction for random noise, are typically expressed within the null hypothesis significance testing (NHST) framework using "*p*-values." This refers to the probability of obtaining test results at least as extreme as the results observed, under the assumption that the null hypothesis is correct. This approach is widespread across the empirical sciences, and many philosophers of science have defended this practice [3, 4].

However, *p*-values and the NHST framework has also received much criticism by other scientists and philosophers [5, 6]. This has become apparent in the study by Bem [7], where empirical data published in an influential psychology journal suggested the presence of precognition (i.e., the ability to predict random future events). This has alarmed psychologists [8] regarding the limitations of the NHST framework. Specifically, it has become apparent that scientists can use questionable research practices (QRPs) to influence the observed results and *p*-values, thereby compromising the validity of scientific findings. Please note that we do not want to suggest that only the NHST framework is sensitive to QRPs (for arguments that it is not when applied properly, see [4]). Indeed, also alternative inferential approaches (e.g., Bayesian hypothesis testing) suffer from limitations due to their sensitivity to misuse and QRPs [9].

Here we provide a short overview of such QRPs and how they can influence the reliability of hypothesis testing and scientific findings. However, note that we do not aim to provide a complete list of such QRPs here. We merely intend to illustrate how different QRPs can result in unreliable findings and incorrect hypothesis tests. For a more exhaustive list of QRPs, see [10].

36.2.1 Not Reporting All Collected Variables

Often, different measurements can be used to test a certain hypothesis. For example, psychologists could test whether social pressure elicits arousal in anxious individuals by measuring skin conductance or heart rate or perhaps both. Indeed, collecting multiple outcome measures is typically seen as good practice to check the generalizability of findings to different measures, as well as help in addressing different research questions within a single study [11]. However, it may be tempting for researchers to selectively report those dependent variables that showed a significant result only and disregard the rest of them. This is especially important since the more tests someone performs within NHST, the higher the changes of a false positive (i.e., the result of random measurement error).

Only presenting the outcome variables that confirm the hypothesis is misleading, as the result can be due to random noise, and the fact that the result was not obtained for the other outcome variables remains obscured. It is therefore commonly recommended by established reporting guidelines to always transparently report on all the collected outcome variables [12].

36.2.2 Failing to Report All Conditions

Much like including different measures in their studies, researchers often also include different conditions in their studies to control for different factors (including both a placebo condition and a wait-list control condition next to the main experimental intervention). When including multiple conditions, results can vary across conditions due to systematic or error variance. For instance, a certain intervention may work well when compared to the wait-list group, but not show a significant effect compared to the placebo group. A researcher can therefore be tempted to only report the comparison between two conditions in which the hypothesis was confirmed, and not report the other conditions. Once again, this compromises the validity of the findings. For instance, it is already well known that interventions tend to have artificially inflated effect sizes when compared to a wait-list control group instead of a placebo-control group [13].

36.2.3 Interim Analyses and Selectively Stopping Data Collection

A third way to influence the results of a study is to collect data and run statistical analyses until a significant result is detected. Due to a fundamental property of the *p*-values, namely, that they tend to decrease with an increasing sample size, additional data collection and uncorrected interim analyses can inflate the chance of a false-positive result. Particularly, *p*-values will always turn out to be significant given a large enough sample of observations [14]. Furthermore, *p*-values tend to fluctuate substantially [15], and it has been argued that the common evidence threshold of $\alpha = .05$ is too liberal and easily results in spurious findings [16]. Given these properties of the *p*-value, *uncorrected* interim analyses and collecting data until the alpha level is crossed will guarantee that a researcher can find a false-positive statistically significant result, thereby greatly increasing the number of spurious results in the literature. Thankfully, there are principled ways to perform interim analyses [17, 18].

36.2.4 Selectively Excluding Data

A fourth way in which results of scientific studies can be compromised is by selectively removing data from the dataset [19, 20]. Due to random noise, there is typically variability in the data of most scientific fields. Without clear pre-specified rules, it is often up to the researchers themselves to decide if certain outliers in the data are due to random error or because of a systematic error that may distort the results. This once again provides an opportunity to capitalize on chance and select those data points that selective support the hypothesis of the researcher, again increasing the chances of finding spurious results.

36.2.5 Changing the Hypothesis After Observing the Results

Another potential way in which the results of a scientific study can be influenced is by adjusting the hypothesis to the observed results. This practice is sometimes referred as "hypothesizing after the results are known" (HARKing) [21]. For instance, a treatment may work in one condition (e.g., low dosage) and not in another (e.g., high dosage). Even though the researchers had initially predicted the opposite pattern, it may be tempting, or it may even happen unintentionally when the hypothesis was not articulated clearly enough beforehand, for the researchers to change their hypothesis. This, however, does not constitute a valid test of a theory because the hypothesis is based on the observed results, rather than specified a priori. Once again, due to random noise in much of the scientific data, this can result in spurious results that are presented as a priori predicted by a flawed theory.

36.2.6 Falsifying and Fabricating Data

Arguably one of the most unethical ways to influence the results of a study is by outright fabrication or falsification of the data to obtain statistically significant "findings" [22]. Though this practice is most likely rare, between 0.3% and 4.9% of researchers self-admit having fabricated or falsified data, and between 5.2% and 33.3% reported personally knowing a colleague who had fabricated or falsified data [23].

36.3 The Consequences of QRPs for Scientific Findings

QRPs can drastically inflate false positives and thus produce unreliable research findings. This has alarmed more and more researcher in recent years. Particularly, several survey studies have shown that scientists self-admit engaging in QRPs [10, 23], and some researchers have raised concerns that as much as 50% or more of the findings in scientific journals are actually false-positive results, in part due to the common (intentional and unintentional) use of QRPs [24, 25]. For example, within the field of psychology, a large-scale replication project of research was only able to replicate 39% of published research findings [26]. Furthermore, a recent study showed that the results of registered reports (a type of journal submission where an article gets accepted before data are collected, merely on the hypotheses and methods to be followed; see below) reported significantly lower percentages of positive results, compared to traditional submissions, casting doubts on how reliable the reported results in the literature are [27]. Results such as these have led several researcher to conclude that psychology and related disciplines are currently suffering from a "replicability crisis" [28, 29]. Nonetheless, similar concerns about the replicability of findings have been raised in other scientific fields, such as cancer research [30], nutrition research [31], and neuroscience [32, 33], indicating that QRPs likely undermine the reliability of scientific findings in many different research areas. This state of (some parts of) the scientific literature is problematic as it has the potential to undermine public trust in scientific research [34] and does not provide stable foundation for further research to be built on.

36.4 Unique Challenges when Using Existing Data

The availability of existing datasets to test hypotheses on can add to the abovementioned problems. Particularly, given that the datasets are already available and that researchers may have preexisting knowledge of their properties, they can use this knowledge to increase their chances to observe statistically significant results, although it is more likely that these results will be spurious or biased. Furthermore, available datasets can often be very large (more than millions of observations) and include many different variables, thereby further increasing the opportunities for using QRPs and finding false-positive results [35, 36].

Additionally, it is more difficult to show with an existing database that a hypothesis was posited prior to looking at the data. When collecting new data, a hypothesis can be publicly announced (for instance, through a preregistration; see the next section) prior to collecting the data. However, when the data are already available, the independence of the hypothesis from the data is more difficult to prove. One exception is when a dataset can only be accessed after approval. In this case, the date of obtaining access to the data can be used to show that a hypothesis was developed independently of the data.

Finally, the widespread availability of datasets to (re-)analyze is relatively new, and there are few guidelines on how to do this correctly. Therefore, it can be argued that establishing guidelines on how to do reliable research and prevent QRPs is particularly important for studies making use of existing data [35, 37].

36.5 Preregistration as a Tool to Protect the Reliability of Scientific Findings

In order to combat these problems with QRPs and unreliable scientific findings, study preregistration has been proposed as a possible solution [38, 39]. In a preregistration document, scientists specify the details regarding their hypotheses, the design of their study, the way in which data are collected, the statistical analysis plan, and the evaluation of the results prior to the execution of the study. This preregistration is typically archived in a (publicly accessible) registration repository prior to conducting the research. By time-stamping this document (i.e., archiving when the preregistration was uploaded to the registry), it can be checked whether the preregistration was available prior to the execution of the study. The preregistration is typically publicly shared prior to the study execution or once the study is accepted, but it could also remain private and only accessible to a selected audience (e.g., co-authors, reviewer, etc.). The idea is that by preregistering important choices in the execution of a study (e.g., sample size, outcome measures, etc.), the flexibility of researchers to (intentionally or unintentionally) influence the results through QRPs is reduced [40] because researchers now specified a plan to follow. If deviations of this plan occur (which can of course happen in a research project), this should then be transparently reported and not presented as an a priori choice.

One of the first areas where the use of preregistrations became widespread is in clinical trials. Many scientific journals within the medical sciences now require clinical trials to be registered in a public repository (e.g., https://www.clinicaltrials.gov/) in order to be considered for publication. More recently, preregistration templates and repositories have also been developed for meta-analyses and systematic reviews (e.g., https://www.crd.york.ac.uk/prospero/) and within the behavioral sciences [41, 42]. For behavioral sciences, the two most common websites for preregistration of a study are osf.io and aspredicted.org, both of which allow researchers to upload their preregistrations for free. An important difference between the two repositories is that although preregistrations in osf.io can stay private for up to 4 years (after which preregistrations become available publicly), in aspredicted.org preregistrations can stay private forever.

36.6 Is Preregistration Necessary for Analyses on Existing Data? Exploratory and Confirmatory Studies

As mentioned above, the typical preregistration document includes a study's hypotheses, methods, planned sample, outcome measures, data transformations, and statistical analyses [42, 43]. In case of the analyses of preexisting data, though, such a complete preregistration is not possible because the study has already been done and the data have already been collected. As such, preregistration of the analyses of preexisting data requires using a different format. However, a first question that needs to be answered is whether the study requires preregistration. It depends on the goal(s) of that study.

First, it may be that a researcher would want to use preexisting data for purely exploratory purposes. In this case, no preregistration is required, given that the researcher has no concrete hypotheses but just performs various tests, attempting to find any interesting data pattern in the data. Although such type of analyses could seem spurious, they are not if they are introduced as such, because the reader can evaluate the provided evidence as stemming from pure exploration. Indeed, exploration is an integral part of the empirical cycle and of scientific discovery [1]. Still, when observing a novel phenomenon in exploratory research, often a follow-up confirmation study is appropriate (i.e., using novel data collection) in order to confirm the novel findings.

For such a "confirmatory study," a researcher could test a clearly specified hypothesis on a dataset that already exists (the European Social

Survey database, the UK Biobank data, etc.). In this case, the hypothesis is specified beforehand, and it is "confirmed" (or disconfirmed) by analyzing independent data (i.e., data that was not used to initially come up with the hypothesis, such as in a prior exploratory study). In a confirmatory study, a preregistration is appropriate and helps ensure that the test of the hypothesis is not tainted by QRPs [44].

We have previously introduced such a template for the preregistration of confirmatory studies on preexisting data [36]. This template consists of ten simple questions, ranging from stating the hypotheses and planned statistical analyses to the provision of clear statements regarding what is already known about the data. This last statement is particularly important because prior knowledge about the data (when dealing with a dataset that has been used before by the same researcher) could limit a study being truly confirmatory (perhaps the researcher already knows that the hypothesized pattern is present in the data). Apart from our own template, also other templates for analyzing secondary data have been recently introduced [37].

36.7 What and How to Preregister: Preregistration Templates and Repositories

Whereas a preregistration is mandatory in many medical journals, other fields are only taking their first steps in applying preregistration for their studies [45]. Therefore, the elements that should be included in the preregistration still differ widely, and, for some types of research designs and in certain scientific fields, no generally accepted standards or templates for preregistration are available as of now [46]. In Table 36.1, we provide a non-exhaustive overview of currently available templates and repositories for some of the most common types of research designs and studies to help guide researchers in choosing an appropriate preregistration format, including preregistration templates for studies using existing data [35–37].

When researchers want to preregister their study, they should complete an appropriate preregistration template for their study and submit this to a relevant repository before conducting their study (or, in the case of analyses on existing data, before inspecting the data and performing the statistical analyses). Furthermore, researchers should refer to this preregistration in the scientific paper resulting from the study. In some journals, preregistration is already required [47], and in other journals, papers with preregistration are designated with open science badges [48].

Finally, a special and powerful version of preregistration are *Registered Reports* [49]. In this format, researchers specify the background of a study, the hypothesis, the design, the sample, the procedure, and the statistical analysis plan, and this is reviewed by a journal *prior* to conducting the study. Once the reviewers and the editor accept this study plan, the authors receive an "in principle acceptance" by the journal, and they can start collecting the data and thereafter submit their final paper. The final paper will again undergo peer review to check whether the study plan was followed. Crucially, the journal will publish the final version of the manuscript, regardless of the obtained results. With this format, QRPs are maximally controlled, and publication of the findings is not based on the direction of the results but merely on the idea and methodology of the study. An updated list of journals offering this publishing format can be found on the Open Science Framework (https://osf.io/rr/).

36.8 Limitations and Critiques of Preregistration

The most common criticism against preregistration is that it does not fit all types of studies, [50], for example, in longitudinal studies, where the final dataset is available only years after the beginning of a study. Given that nowadays statistical methodologies develop rapidly, it is probable that by the end of a study, a better method for addressing the research question will be available. However, since this analysis was not

Table 36.1 Non-exhaustive review of preregistration templates and repositories for different types of studies

Type of study	Preregistration templates	Possible public repositories
Qualitative studies	Qualitative Preregistration template [46]	https://osf.io/
Quantitative behavioral research	As Predicted template (https://osf.io/fnsb6/) Pre-Registration in Social Psychology [42]; https://osf.io/ce3hr/)	https://osf.io/ https://aspredicted.org/
Analyses on existing data	Mertens and Krypotos [36] template (https://osf.io/3tbwc/) Weston et al. [35] template (https://osf.io/x4gzt/)	https://osf.io/
Randomized controlled trials	Use the protocol registration and results system (see www.clinicaltrials.gov)	https://www.clinicaltrials.gov/ https://www.clinicaltrialsregister.eu/
Systematic reviews/ meta-analyses	Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P)	https://www.crd.york.ac.uk/prospero/ https://srdr.ahrq.gov/

Note: A more exhaustive list of preregistration templates can be found on the Open Science Framework: https://osf.io/zab38/wiki/home/

preregistered, researchers may worry that it is not allowed to perform it.

This idea stems from the common misconception that after the submission of a study's preregistration, researchers' hands are tied, and any deviation from the plan should be interpreted as engaging in QRPs. In our view, however, this is an unfortunate misrepresentation and misinterpretation of the goals and implementation of preregistrations. The preregistration is a plan, and a plan can be changed or updated [51]. As long this is done in a transparent manner, by updating the preregistration template before the data are inspected, with the authors describing the changes in their preregistration or by listing the deviations from the preregistration in the final paper, there is no reason to hang on to the choices made in the preregistration and for the authors to be accused of QRPs.

Another common argument is that preregistration cannot really protect from QRPs, simply because someone could preregister a study after the data have been inspected, or preregister multiple studies, keep the preregistrations private, and then only release the preregistration that better fits the direction of the results. Alternatively, and less dramatically, preregistrations could simply not be clearly specified enough and/or not followed by researchers and thereby not really protect against QRPs [52, 53].

We agree that indeed a study's preregistration is not a tool that can guarantee a 100% safeguard against QRPs and a preregistered study is not necessarily a good study. Still, we caution against throwing the baby out with the bathwater. Given that QRPs are fairly common and often happen unintentionally [23, 54], preregistration is a valuable tool to encourage researchers to be transparent in their choices. Furthermore, even if an intervention is not 100% effective, it can nonetheless be a useful tool to reduce, though perhaps not eliminate, (unintentional) QRPs.

36.9 Closing Remarks and Future Perspectives

We focused mostly on preregistration as a tool to reduce (unintended) QRPs and increase transparency and did not evaluate preregistration in its ability to test the severeness of a test, a topic that is most relevant for the field of philosophy of science [44].

Although the idea of preregistering studies has been around since the 1960s, researchers have only relatively recently started using preregistration for research in different areas (e.g., psychology, philosophy, social sciences). As such, for these fields, preregistration is a relatively new tool, and it is still being further developed and

evaluated [45, 46]. Given the relatively quick developments in the field of open science in psychology and beyond, we anticipate that a study's preregistration will become more common for different types of research designs and scientific disciplines. This will be a good step towards promoting more transparency in our research. Such transparency is important for science consumers in evaluating scientific research and for having trust in scientific research findings [34].

Preregistration is not a catchall solution for other problems that may exists in different scientific fields, such as vague theories or poor external validity [55]. Finally, it is important that preregistration be sufficiently specific and actually followed. Indeed, a number of recent studies found that preregistrations are not always followed carefully, and this is often not reported transparently [52, 53]. In these cases, the usefulness of preregistration to protect against QRPs is obviously diminished, and papers may undeservedly receive credit for good practices that were not actually adhered to. That said, a preregistration does not and should not prevent researchers from choosing the optimal statistical models and data points for their research aims, and researchers should always be allowed to explore their data to discover new patterns and findings (provided that this is reported as such).

The wider acceptability of preregistration in the community is likely going to take more effort as it calls for a wider change in the scientific culture and the current way of doing science. Typically most of the important decisions (e.g., the exact hypothesis, the statistical model, data exclusions, etc.) are now commonly taken during or at the end of the study, while this should preferably be done beforehand. Furthermore, most researchers are under significant pressure to publish articles and journals and often prefer publishing positive results. Flexibility in the specification of the details of a study can help researchers find such positive results and thus publish more easily [54, 56]. As such, many researchers are still disincentivized from adopting the practice of preregistering their study. Nonetheless, given that preregistration provides important advantages to the transparency and reliability of scientific research, we expect that more funders, universities, and journals will require researchers to preregister their studies in the future.

References

1. De Groot AD (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 L. J. van der Maas]. Acta Psychol 148:188–194
2. Popper K (1959) The logic of scientific discovery. Hutchinson, London
3. Mogie M (2004) In support of null hypothesis significance testing. Proc R Soc Lond Ser B Biol Sci 7:271. https://royalsocietypublishing.org/doi/10.1098/rsbl.2003.0105
4. Lakens D (2021) The practical alternative to the p value is the correctly used p value. Perspect Psychol Sci 16:639–648. http://journals.sagepub.com/doi/10.1177/1745691620958012
5. Carver R (1978) The case against statistical significance testing. Harv Educ Rev 48(3):378–399. https://meridian.allenpress.com/her/article/48/3/378/21418/The-Case-Against-Statistical-Significance-Testing
6. McShane BB, Gal D, Gelman A, Robert C, Tackett JL (2019) Abandon statistical significance. Am Stat 73:235–245. https://www.tandfonline.com/doi/full/10.1080/00031305.2018.1527253
7. Bem DJ (2011) Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. J Pers Soc Psychol 100(3):407–425. http://doi.apa.org/getdoi.cfm?doi=10.1037/a0021524
8. Wagenmakers E-J, Wetzels R, Borsboom D, van der Maas HLJ (2011) Why psychologists must change the way they analyze their data: the case of psi: Comment on Bem (2011). J Pers Soc Psychol 100(3):426–432. http://doi.apa.org/getdoi.cfm?doi=10.1037/a0022790
9. Tendeiro JN, Kiers HAL (2019) A review of issues about null hypothesis Bayesian testing. Psychol Methods 24(6):774–795. http://doi.apa.org/getdoi.cfm?doi=10.1037/met0000221
10. John LK, Loewenstein G, Prelec D (2012) Measuring the prevalence of questionable research practices with incentives for truth telling. Psychol Sci 23(5):524–532. http://journals.sagepub.com/doi/10.1177/0956797611430953
11. LoBue V, Reider LB, Kim E, Burris JL, Oleas DS, Buss KA et al (2020) The importance of using multiple outcome measures in infant research. Infancy 25(4):420–437. https://onlinelibrary.wiley.com/doi/abs/10.1111/infa.12339

12. Schulz KF, Altman DG, Moher D (2010) CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Trials 11(1):32. https://trialsjournal.biomedcentral.com/articles/10.1186/1745-6215-11-32
13. Cuijpers P, Cristea IA (2016) How to prove that your therapy is effective, even when it is not: a guideline. Epidemiol Psychiatr Sci 25(5):428–435. https://www.cambridge.org/core/product/identifier/S2045796015000864/type/journal_article
14. Wagenmakers E-J (2007) A practical solution to the pervasive problems of p values. Psychon Bull Rev 14(5):779–804. http://www.springerlink.com/index/10.3758/BF03194105
15. Cumming G (2014) The new statistics. Psychol Sci 25(1):7–29. http://journals.sagepub.com/doi/10.1177/0956797613504966
16. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, Berk R et al (2018) Redefine statistical significance. Nat Hum Behav 2(1):6–10. http://www.nature.com/articles/s41562-017-0189-z
17. Lakens D (2014) Performing high-powered studies efficiently with sequential analyses. Eur J Soc Psychol 44(7):701–710. http://doi.wiley.com/10.1002/ejsp.2023
18. Schönbrodt FD, Wagenmakers E-J, Zehetleitner M, Perugini M (2017) Sequential hypothesis testing with Bayes factors: efficiently testing mean differences. Psychol Methods 22(2):322–339. http://doi.apa.org/getdoi.cfm?doi=10.1037/met0000061
19. Lonsdorf TB, Klingelhöfer-Jens M, Andreatta M, Beckers T, Chalkia A, Gerlicher A et al (2019) Navigating the garden of forking paths for data exclusions in fear conditioning. elife 8:e52465
20. Morís Fernández L, Vadillo MA (2020) Flexibility in reaction time analysis: many roads to a false positive? R Soc Open Sci 7(2):190831. https://royalsocietypublishing.org/doi/10.1098/rsos.190831
21. Kerr NL (1998) HARKing: hypothesizing after the results are known. Personal Soc Psychol Rev 2(3):196–217
22. Neuroskeptic (2012) The nine circles of scientific hell. Perspect Psychol Sci 7(6):643–644. http://journals.sagepub.com/doi/10.1177/1745691612459519
23. Fanelli D (2009) How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS One 4(5):e5738. https://dx.plos.org/10.1371/journal.pone.0005738
24. Simmons JP, Nelson LD, Simonsohn U (2011) False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol Sci 22(11):1359–1366. http://journals.sagepub.com/doi/10.1177/0956797611417632
25. JPA I (2005) Why most published research findings are false. PLoS Med 2(8):e124. https://dx.plos.org/10.1371/journal.pmed.0020124
26. Open Science Collaboration (2015) Estimating the reproducibility of psychological science. Science 349(6251):aac4716. http://www.sciencemag.org/cgi/doi/10.1126/science.aac4716
27. Scheel AM, Schijen M, Lakens D (2021) An excess of positive results: comparing the standard psychology literature with registered reports. Adv Methods Pract Psychol Sci 4:25152459211007467
28. Pashler H, Wagenmakers E (2012) Editors' introduction to the special section on replicability in psychological science. Perspect Psychol Sci 7(6):528–530. http://journals.sagepub.com/doi/10.1177/1745691612465253
29. Tackett JL, Lilienfeld SO, Patrick CJ, Johnson SL, Krueger RF, Miller JD et al (2017) It's time to broaden the replicability conversation: thoughts for and from clinical psychological science. Perspect Psychol Sci 12(5):742–756. http://journals.sagepub.com/doi/10.1177/1745691617690042
30. Wen H, Wang H-Y, He X, Wu C-I (2018) On the low reproducibility of cancer studies. Natl Sci Rev 5(5):619–624. https://academic.oup.com/nsr/article/5/5/619/4835582
31. Sorkin BC, Kuszak AJ, Williamson JS, Hopp DC, Betz JM (2016) The challenge of reproducibility and accuracy in nutrition research: resources and pitfalls. Adv Nutr 7(2):383–389. https://academic.oup.com/advances/article/7/2/383/4558081
32. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ et al (2013) Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 14:365–376
33. Botvinik-Nezer R, Holzmeister F, Camerer CF, Dreber A, Huber J, Johannesson M et al (2020) Variability in the analysis of a single neuroimaging dataset by many teams. Nature 582(7810):84–88. http://www.nature.com/articles/s41586-020-2314-9
34. Wingen T, Berkessel JB, Englich B (2020) No replication, no trust? How low replicability influences trust in psychology. Soc Psychol Personal Sci 11(4):454–463. http://journals.sagepub.com/doi/10.1177/1948550619877412
35. Weston SJ, Ritchie SJ, Rohrer JM, Przybylski AK (2019) Recommendations for increasing the transparency of analysis of preexisting data sets. Adv Methods Pract Psychol Sci 2(3):214–227. http://journals.sagepub.com/doi/10.1177/2515245919848684
36. Mertens G, Krypotos A-M (2019) Preregistration of analyses of preexisting data. Psychol Belg 59(1):338–352. http://www.psychologicabelgica.com/articles/10.5334/pb.493/
37. van den Akker O, Weston SJ, Campbell L, Chopik WJ, Damian RI, Davis-Kean PE et al (2019) Preregistration of secondary data analysis: a template and tutorial. PsyArXiv. https://doi.org/10.31234/osf.io/hvfmr
38. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N et al (2017) A manifesto for reproducible science. Nat Hum Behav 1(1):0021. https://doi.org/10.1038/s41562-016-0021
39. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ et al (2015) Promoting an open

research culture. Science 348(6242):1422–1425. http://www.sciencemag.org/cgi/doi/10.1126/science.aab2374
40. Nosek BA, Beck ED, Campbell L, Flake JK, Hardwicke TE, Mellor DT et al (2019) Preregistration is hard, and worthwhile. Trends Cogn Sci 23(10):815–818. https://linkinghub.elsevier.com/retrieve/pii/S1364661319301846
41. Krypotos A-M, Klugkist I, Mertens G, Engelhard IM (2019) A step-by-step guide on preregistration and effective data sharing for psychopathology research. J Abnorm Psychol 128(6):517–527
42. van 't Veer AE, Giner-Sorolla R (2016) Pre-registration in social psychology—a discussion and suggested template. J Exp Soc Psychol 67:2–12. https://doi.org/10.1016/j.jesp.2016.03.004
43. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT (2018) The preregistration revolution. Proc Natl Acad Sci 115(11):2600–2606. http://www.pnas.org/lookup/doi/10.1073/pnas.1708274114
44. Lakens D (2019) The value of preregistration for psychological science: a conceptual analysis. Japanese Psychol Rev 62(3):221–230
45. Polonioli A, Vega-Mendoza M, Blankinship B, Carmel D (2021) Reporting in experimental philosophy: current standards and recommendations for future practice. Rev Philos Psychol 12:49–73. http://link.springer.com/10.1007/s13164-018-0414-3
46. Haven TL, Errington TM, Gleditsch KS, van Grootel L, Jacobs AM, Kern FG et al (2020) Preregistering qualitative research: a Delphi Study. Int J Qual Methods 1(19):160940692097641. http://journals.sagepub.com/doi/10.1177/1609406920976417
47. DeAngelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R et al (2005) Clinical trial registration. Arch Dermatol 141(1):76–77. http://archderm.jamanetwork.com/article.aspx?doi=10.1001/archderm.141.1.76
48. Kidwell MC, Lazarević LB, Baranski E, Hardwicke TE, Piechowski S, Falkenberg L-S et al (2016) Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency. PLOS Biol 14(5):e1002456. https://dx.plos.org/10.1371/journal.pbio.1002456
49. Chambers CD (2013) Registered reports: a new publishing initiative at Cortex. Cortex 49(3):609–610. https://doi.org/10.1016/j.cortex.2012.12.016
50. Pham MT, Oh TT (2021) Preregistration is neither sufficient nor necessary for good science. J Consum Psychol 31(1):163–176. https://onlinelibrary.wiley.com/doi/10.1002/jcpy.1209
51. DeHaven AC (2017) Preregistration: a plan, not a prison. COS blog [Internet]. https://www.cos.io/blog/preregistration-plan-not-prison
52. Claesen A, Gomes S, Tuerlinckx F, Vanpaemel W (2019) Preregistration: comparing dream to reality. PsyArXiv
53. Bakker M, Veldkamp CLS, van Assen MALM, Crompvoets EAV, Ong HH, Nosek BA et al (2020) Ensuring the quality and specificity of preregistrations. PLoS Biol 18(12):e3000937. https://dx.plos.org/10.1371/journal.pbio.3000937
54. Grant DB, Kovács G, Spens K (2018) Questionable research practices in academia: antecedents and consequences. Eur Bus Rev 30(2):101–127. https://www.emerald.com/insight/content/doi/10.1108/EBR-12-2016-0155/full/html
55. Szollosi A, Kellen D, Navarro DJ, Shiffrin R, van Rooij I, Van Zandt T et al (2020) Is preregistration worthwhile? Trends Cogn Sci 24(2):94–95. https://linkinghub.elsevier.com/retrieve/pii/S1364661319302852
56. Fanelli D (2010) Do pressures to publish increase scientists' bias? An empirical support from US States data. PLoS One 5(4):e10271. https://dx.plos.org/10.1371/journal.pone.0010271