REVIEW

Reducing bias, increasing transparency, and calibrating confidence with preregistration

Tom E. Hardwicke* & Eric-Jan Wagenmakers#

Department of Psychology, University of Amsterdam

Flexibility in the design, analysis, and interpretation of scientific studies creates a multiplicity of possible research outcomes. Scientists are granted considerable latitude to selectively use and report the hypotheses, variables, and analyses, that create the most positive, coherent, and attractive story, whilst suppressing those that are negative or inconvenient. This creates a risk of bias that can lead to scientists fooling themselves and fooling others. Preregistration involves declaring a research plan (e.g., hypotheses, design, and statistical analyses) in a public registry before the research outcomes are known. Preregistration (1) reduces the risk of bias by encouraging outcome-independent decision-making; and (2) increases transparency, enabling others to assess the risk of bias and calibrate their confidence in research outcomes. In this article, we briefly review the historical evolution of preregistration in medicine, psychology, and other domains, clarify its pragmatic functions, discuss relevant meta-research, and provide recommendations for scientists and journal editors.

*hardwickete@gmail.com ORCID: 0000-0001-9485-4952

#ej.wagenmakers@gmail.com ORCID: 0000-0003-1596-1034

Introduction

Scientific research is performed by fallible humans. Our cognitive limitations and self-interested motivations can infuse bias into the research process, undermining the production of reliable knowledge (Box 1)[1–3]. The contemporary scientific ecosystem, in which research is funded, conducted, and disseminated, actively perpetuates bias by primarily rewarding scientists based on the nature of research outcomes over the quality of research methods[4–6]. Biased research is wasteful[7], perpetuates falsehoods[8], and can lead to applied interventions that are premature, ineffective, or even harmful[9,10]. There is concern in many fields that scientists are fooling themselves, and fooling each other, at a frequency that is unacceptably high[11–14].

These problems notwithstanding, the obvious success of science — from vaccines to Mars rovers — is a testament to the potential of scientific methods. A key contributor to this success is the ongoing development of tools that help scientists to learn about the world without fooling themselves[15]. For example, statistical tools help to differentiate signals from noise, randomisation helps to isolate causal mechanisms, and placebos help to control for participant reactivity. There is growing interest in a tool called *preregistration* which involves declaring a research plan (e.g., aims/hypotheses, design, and statistical analyses) in a public registry before the research outcomes are known[16,17]. Preregistration (1) reduces the risk of bias by encouraging outcome-independent decision-making; and (2) increases transparency, enabling others to assess the risk of bias and calibrate their confidence in research outcomes. In this article, we briefly review the historical evolution of preregistration in medicine, psychology, and other domains, clarify its pragmatic functions, discuss relevant meta-research, and provide recommendations for scientists and journal editors.

Box 1. Human after all: Cognitive bias and skewed incentives

The storybook image of the scientist as an objective, rational, and dispassionate arbiter of truth is apparently pervasive amongst both lay people and scientists themselves[18]. But despite these pretensions, scientists have egos, career ambitions, and rent to pay[19]. Any intrinsic motivation toward the pure pursuit of knowledge may be undermined by extrinsic motivations towards producing the most fundable or publishable research. Currently, the allocation of funding, awards, and publication prestige predominantly rewards scientists based on their research outcomes being impressive over being right[4–6]. Typically, this manifests as a preference for novel, positive, and 'statistically significant' outcomes over incremental, negative, or null outcomes[20,21]. There is additional pressure to produce articles with concise, coherent, and compelling narratives, encouraging selective reporting of research methods and results in order to hide the messy realities of scientific inquiry beneath a veneer of artificial perfection[22,23].

Psychological research has documented an array of cognitive biases that can create systematic errors in reasoning and belief updating. Even appropriate incentives, expertise, and good intentions may not be sufficient to overcome the influence of cognitive biases and there is some (though limited) evidence of their impact in scientific contexts[24–26]. Human fallibility highlights the need for intellectual humility and transparency to enable our peers to properly evaluate our research[27]. Relevant examples of cognitive bias include: *confirmation bias* — a tendency to preferentially seek out, evaluate, and recall information that supports one's existing beliefs[28]; *motivated reasoning* — generating post-hoc rationalizations that frame previous decisions in a favourable light[29]; *hindsight bias* — a tendency to think past events had a higher likelihood of occurring relative to our actual prior predictions of the same events ("I knew it all along")[30,31]; *apophenia* — a tendency to identify seemingly meaningful patterns in random data[32]; and the *bias blind spot* — a lack of awareness about how our own decisions are influenced by bias[33].

A brief history of preregistration

Philosophers and methodologists have long debated the epistemic merits of 'predesignating' hypotheses before confronting them with evidence[34–37]. The practical idea of preregistering a study in a registry, or with a journal, was contemplated intermittently in the social and behavioural sciences during the 1960s-1980s[38,39]. However, aside from a remarkable forebear of the contemporary *Registered Reports* format (Box 2) offered at the *European Journal of Parapsychology* between 1976 and

1992[40], early preregistration pondering lacked the necessary cultural impetus and technological infrastructure to initiate substantive changes in research practice[41].

In medicine, concern about publication bias prompted proposals for an international clinical trial registry in the 1980s[42]. Clinical trial registration became a requirement for publication in major medical journals in 2004[43] and in subsequent years became a legal requirement in some jurisdictions[44]. In the most recent amendment of the Declaration of Helsinki, an ethical mandate for the registration of every research study involving human participants was introduced[45]. However, despite registration mandates, compliance remains a major problem, with many trials registered retrospectively or affected by selective reporting[46,47]. Additionally, the motivation and infrastructure of clinical trial registration places less emphasis on detailed prespecification of statistical analysis plans relative to other non-medical fields and in practice many trial registrations do not include detailed analysis plans[48–50] leaving scope for bias[51]. Contemporary discussions and registry infrastructure often assume that a preregistration can contain information about a study's aims/hypotheses, design, and analyses, a framework we will also adhere to in this review. Preregistration of other research designs, such as observational studies, has been advocated[52] and contested[53], and is less common[54]. Nevertheless, editors at the BMJ recently described preregistration as "The single most valuable tool we have to ensure unbiased reporting of research studies"[55].

As a multidisciplinary 'crisis of confidence' in research has unfolded over the last decade[11–14], interest in preregistration beyond medicine has increased dramatically[17], prompting debate in psychology[56–60], political science[61,62], economics[63,64], cognitive modelling[65,66], neuroimaging[67], secondary data analysis[68], and qualitative research[69] (Supplementary Information A). Several domain-specific registries have been established (Supplementary Information B). Registry entries[60,70,71] and survey evidence[72] signal increasing adoption of preregistration; however, prevalence estimates derived from manual inspection of randomly sampled empirical articles suggest that it remains uncommon overall in biomedicine[73], psychology[74], and the social sciences[75]. The contemporary Registered Reports format (Box 2) was introduced in 2013[76] and is

currently offered by more than 300 journals, including *Cortex*, *Royal Society Open Science*, and *Nature Human Behaviour*[77,78].

Box 2. Complementary tools and extensions

Several tools and extensions may enhance preregistration[79,80]. As these tools can introduce researcher degrees of freedom and do not protect against selective reporting, they should ideally complement rather than replace preregistration.

Robustness checks

Whilst preregistration aims to constrain researcher degrees of freedom, robustness checks directly exploit them in order to evaluate their impact. Traditional sensitivity analyses may evaluate a few justifiable options for a single research decision[81]; however, recent approaches, variously known as "multiverse analysis"[82], "vibration of effects"[83], "specification curve"[84], or "multimodel analysis"[85], systematically assess the factorial intersection of multiple choices for multiple decisions, potentially resulting in tens of thousands of unique analysis specifications[86–88]. This is akin to simultaneously examining multiple cells (research outcomes) in the array depicted in Figure 1, rather than a single prespecified cell.

Some have argued that systematic robustness checks render preregistration redundant[59,89]. However, the subjective choice of which specifications to examine or report[80,90] means that robustness checks can introduce researcher degrees of freedom, creating an opportunity for selective reporting, and thereby increasing the risk of bias. Researchers can have the best of both worlds by preregistering their robustness checks.

Blind analysis

Issues arising during data collection such as attrition, missing data, randomisation failures, or unexpected data distributions may invalidate planned analyses. Blind analysis disguises information related to outcomes (e.g., by adding noise or shuffling variables) allowing the data to be inspected whilst ensuring decision-making remains outcome-independent[91,92]. Blind analysis is used in physics to address concerns about bias introduced by outcome-dependent analyses[93]. Blind analysis requires some technical expertise and can introduce bias if poorly implemented. Additionally, blind analysis does not prevent selective reporting, so should ideally be used in conjunction with preregistration.

Hold-out samples

Splitting a dataset can enable exploratory (outcome-dependent) analyses in a 'training' sample followed by confirmatory analyses in a 'test' or 'hold-out' sample[94]. This approach requires a large sample size as splitting the data reduces statistical power. Preregistering the analyses intended for the hold-out sample ensures they are truly confirmatory (outcome-independent).

Preregistration of analysis scripts based on simulated data

It requires some experience and foresight to anticipate the details to include in an analysis plan[56,95] and it can be difficult to communicate analysis specifications in prose[96,97]. This can be addressed by preregistering analysis scripts prepared using simulated data[51].

Standard Operating Procedures

Maintaining a living document of default research decisions that is co-registered with each study could enhance preregistration efficiency[98].

Open lab notebooks

Open lab notebooks could improve transparency throughout a research project[99] and help track departures from the preregistration. Preregistration is similar to sharing the pages of your lab notebook which outline the study plan.

Registered Reports.

Registered Reports are a journal article format that offers in-principle acceptance for publication before studies begin based on peer review of a prespecified study protocol[39,76]. This radical departure from traditional publication practices promises the benefits of preregistration with enhanced protection against publication bias. Registered Reports may be most suitable for more confirmatory research when most research decisions can be anticipated in advance[100]. Related tools that could also mitigate publication bias involve combining standard preregistration with post-study results-blind peer review[101] and encouraging results reporting in study registries (a mandate for some clinical trials, often ignored in practice[102,103]); however, the time investment and public accountability involved in Registered Reports may offer greater motivation for authors to report results even if they do not reflect their preferred outcome.

What is the problem preregistration is trying to address?

Multiplicity, outcome-dependent decision-making, and bias

There is no single 'correct' way to design, analyze, and interpret a scientific study. Most research decisions have several justifiable options (*degrees of freedom*), giving researchers considerable control over how evidence is generated (*evidentiary degrees of freedom*) and interpreted (*interpretive degrees of freedom*; Box 3)[104–106]. For example, researchers must decide which variables to measure, when to stop data collection, how to handle potential outliers, and which hypotheses to test. The factorial combination of these many evidentiary and interpretive decisions gives rise to a multiplicity of possible research outcomes (Figure 1). The number of possible outcomes is potentially very large — for example, in a highly simplified hypothetical case involving only five research decisions, each with 5 justifiable options, there would be 3125 (5^5) unique outcomes. Researcher degrees of freedom may be naturally constrained to some extent by strong theory[62,89,107], community norms[14], or in replication studies, though these constraints may be more implicit than explicit, leaving room for flexible decision-making.

Why is this a problem? Scientific observations typically consist of both noise (random variation unique to the current sample) and signal (regularities that will reoccur in other samples). Consequently, some evidence will inevitably be misleading (e.g., inflated effect sizes, exaggerated evidence, or false positives)[14,108]. The problem is further exacerbated in situations involving small effect sizes, high variation, and large measurement errors, as these all reduce the signal-to-noise ratio[14,109]. Moreover, for any given evidence, there may be a large number of relevant explanatory theories[36,110] and for any given theory one may be able to derive several specific hypotheses, especially when theories are ambiguously specified and flexible[111–113]. Consequently, scientists are often operating in an environment in which they can easily mistake ephemeral coincidences (noise) for genuine discoveries (signals) and retroactively construct or select from a large range of plausible interpretations.

When researchers make design, analysis, and interpretive decisions with knowledge of research outcomes (*outcome-dependent* decision-making), they may act on inappropriate incentives (Box 1) and make choices which skew the outcomes towards

larger, positive, 'statistically significant' effects that align with their preferred hypothesis. In doing so, they increasing their likelihood of encountering misleading evidence and 'capitalizing on chance'. Thus, a multiplicity of possible research outcomes, combined with outcome- dependent decision-making, creates a *risk of bias* (systematic deviation of results and interpretation from the truth)[114–116].

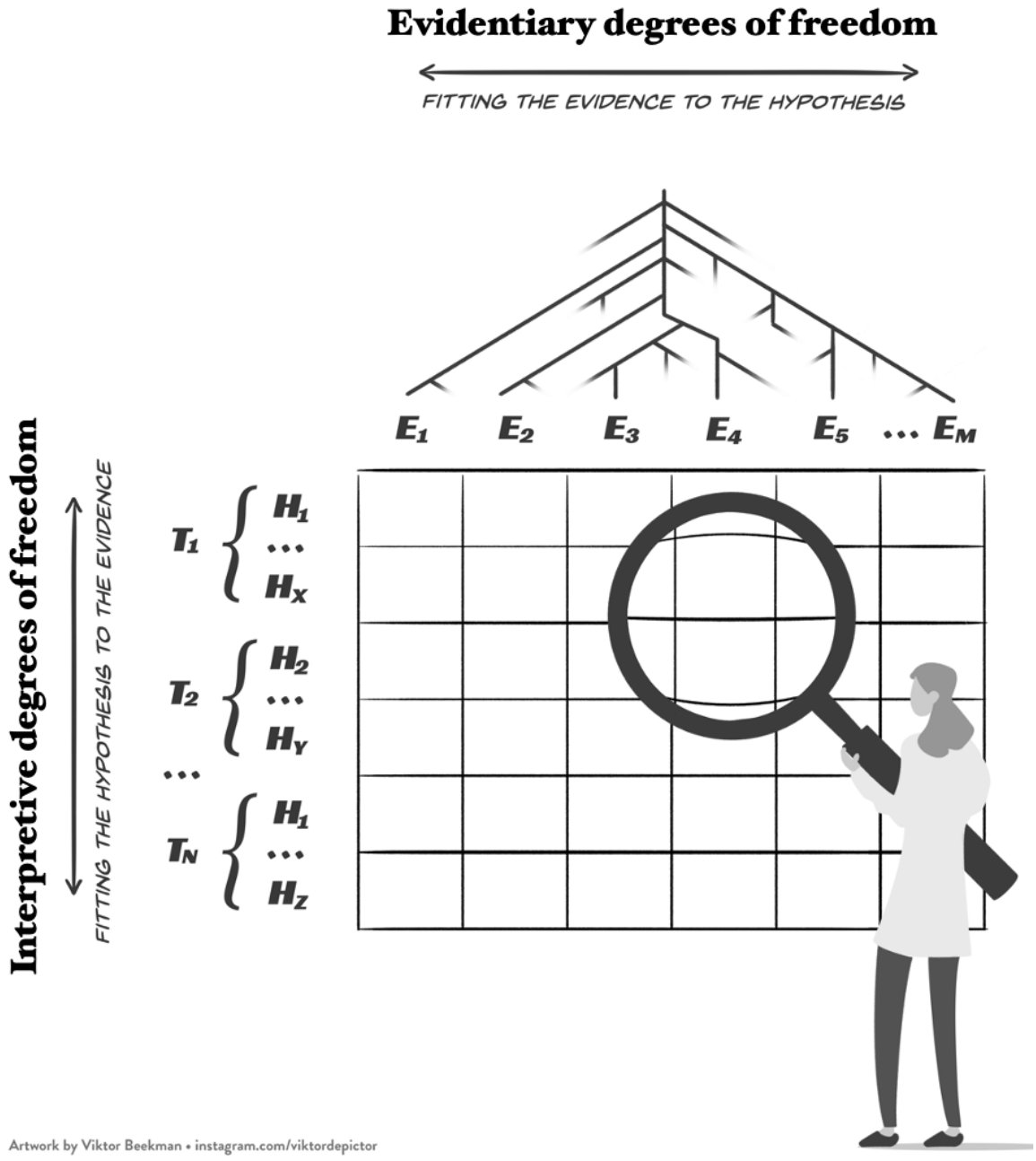

Figure 1. A researcher exploring an array of study outcomes. The horizontal axis illustrates a highly simplified 'garden of forking paths': the many (potentially tens of thousands) justifiable design and analysis specifications that researchers can use to generate evidence (E). The vertical axis illustrates that there may be several relevant

theories (T), and hypotheses (H) derived from those theories, which could be constructed or selected and then confronted with the evidence. An unconstrained researcher can simultaneously exploit their evidentiary degrees of freedom and interpretive degrees of freedom to fit evidence to hypotheses and fit hypotheses to evidence in order to arrive at a research outcome that aligns with their preferences. *Adapted from https://tinyurl.com/3vnpr982 under a CC-BY license.*

Box 3. Researcher degrees of freedom

Researchers often have many justifiable choices when they make decisions about how to design studies and analyze data in order to generate evidence (*evidentiary degrees of freedom*) and interpret that evidence (*interpretive degrees of freedom*; Figure 1)[104–106].

Evidentiary degrees of freedom

Evidentiary degrees of freedom refer to the design and analysis decisions that determine how data are constructed or selected (i.e., defined, collected, processed, filtered, and summarised) in order to transform them into evidence. Examples include decisions about outlier removal[117], designation of primary/secondary measured variables[118], multiplicity corrections[119], adjustment for covariates[51], handling missing data[120], subgroup analyses[121], data collection stopping rules based on interim results[122], and selection of measured variables[123]. Exploiting evidentiary degrees of freedom (intentionally or unintentionally) in a manner that biases the evidence is variously known as 'cooking' the data[124], 'cherry picking'[125], 'p-hacking'[60,106], and 'specification searching'[104]. As noted by Barber (p. 20)[19], "When not planned beforehand, data analysis can approximate a projective technique, such as the Rorschach, because the investigator can project on the data his own expectancies, desires, or biases and can pull out of the data almost any 'finding' he may desire."

Interpretive degrees of freedom

Interpretive degrees of freedom refer to decisions about the construction or selection of general theories and which specific hypotheses derived from those theories are to be confronted with (and thus potentially confirmed or falsified by) the evidence. The extent to which 'hypothesising after the results are known' (HARKing)[126] is problematic has been contested[59,89,127]. An enduring debate in philosophy of science has contemplated whether the prediction of evidence is epistemically superior to the accommodation of evidence[128,129]. It is generally agreed that it is not the temporal order of the hypothesis and the evidence that matters per se, but whether the

construction or selection of hypotheses was independent of the evidence (prediction) or dependent on the evidence (accommodation). By some accounts, even this distinction is irrelevant, because all that matters for theory evaluation is the relationship between the hypothesis and the evidence[35,36,130]. However, others see epistemic or pragmatic value in prediction over accommodation because it reduces the likelihood of overfitting[131], prevents 'fudging' (i.e., tweaking a hypothesis to fit the evidence)[132], increases test severity[133–135], and implies that the hypothesis was generated by a reliable method[136] or is more likely to be true[137]. Others have also argued that scientific progress depends primarily on the falsification of theories, which can only be achieved by testing predictions[138,139]. HARKing could also introduce bias through the preferential construction or selection of hypotheses that successfully fit the data whilst neglecting hypotheses that do not. This is a form of confirmation bias (Box 1). HARKing may also entail a risk of 'double counting' the evidence if a researcher loses track of their original prior (see main text). Further discussion of the prediction-accommodation issue is beyond the scope of this paper, but we point the reader to contemporary reviews[128,129].

Importantly, when decision-making is outcome-dependent, bias can occur even if a researcher does not explicitly evaluate different decision options. For example, imagine an analysis decision for which there are three justifiable choices. A researcher tries one of these options and observes that the outcome of the analysis is statistically significant. Had the outcome been different, the researcher may have also evaluated other methods; however, they are less likely to do so if they are satisfied with the outcome of the first method. Because the researcher's decision is influenced by the outcome, they are more likely to have obtained exaggerated evidence (i.e., there is an increased risk of bias). This is also an example of 'regression to the mean'[116]. When a particular measurement is selected because it 'stands out' or crosses some threshold (e.g., 'statistical significance'), then it is more likely to provide an exaggerated (i.e., biased) estimate that decreases upon subsequent measurement (i.e., it regresses to the mean). It is often easy to convince oneself after the fact, that there was a strong outcome-independent rationale for making a particular decision (see 'motivated reasoning', Box 1). Thus, a researcher engaged in outcome-dependent decision-making need not be explicitly hunting for a particular kind of outcome to be misled by multiplicity.

Various statistical philosophies have principled ways of addressing multiplicity, however these are intended to address the problem of testing multiple hypotheses

rather than outcome-dependent decision-making. Additionally, the proper functioning of these statistical safeguards depends on knowing the extent of multiplicity in order to adjust for it (either through frequentist or objective Bayesian procedures[140,141]) or that researchers' specific beliefs about the plausibility of each hypothesis being considered are explicitly encoded in their priors (in subjective Bayesian procedures[142,143]). If a researcher is engaged in outcome-dependent decision-making, it can be difficult to keep track of the multiplicity they are exposed to — there is a multiple comparisons problem with the number of comparisons unknown[34]. Due to hindsight bias and the limits of mental time travel, it is also difficult to retroactively specify outcome-independent prior probabilities once outcomes have been observed (Box 1). This creates a risk of double counting the evidence in subjective Bayesian analyses[144] because there is an initial informal 'update' when a researcher constructs or selects a hypothesis that seems relevant given the evidence, followed by a formal update when the hypothesis is confronted with the same evidence. In sum, navigating researcher degrees of freedom in an outcome-dependent manner increases the risk of bias and can compromise common statistical safeguards such that inferential tools like p-values and Bayes factors cannot be taken at face value[34,144,145].

How bad is it?

Several lines of evidence document the risk of bias inherent in outcome-dependent decision-making. The problem has been illustrated multiple times with simulations and empirical demonstrations[51,106,122] (for review see ref[146]). For example, researchers have deliberately exploited their degrees of freedom to report that Ontario residents with the star sign 'Leo' have a significantly higher risk of gastrointestinal hemorrhage[147]; participants who listen to "When I'm 64" by the Beatles become physically younger[106]; and 'brain activity' can be detected in a dead Atlantic Salmon during a cognitive perspective-taking task[148]. Beyond deliberately absurd demonstrations, systematic robustness checks (Box 2) suggest that examining outcomes across a large range of analysis specifications can result in considerable variation in outcomes, or 'vibration of effects'[82–84]. Additionally, several "Many Analyst" projects have found variation in analytic approaches and resulting outcomes when different research teams tackle the same research question with the same data[149–151]. This suggests that, at least in principle,

researcher degrees of freedom and outcome-dependent decision-making can create opportunities for bias to distort research outcomes (for review see ref[87]).

To what extent are researcher degrees of freedom exploited in practice and what impact does this have? In surveys, researchers have admitted to exploiting their researcher degrees of freedom (for example, selective outcome reporting) and consider it likely that their colleagues do too[152–154]. One survey of biostatisticians found that they frequently received inappropriate requests from researchers asking them to find a way to produce 'better' outcomes[155]. Signatures of bias have also been detected in the published literature; in particular, the suspicious juxtaposition of overwhelmingly statistically significant results[156–158] and inadequate statistical power[159,160]. The results of multiple large-scale replication studies consistently imply that many published results are either exaggerated by bias or entirely spurious[161–163] (for review see ref[71]). Signatures of bias can also be detected using meta-analytic tools[164], though these cannot always effectively measure or correct for bias[165,166]. Several studies have identified direct evidence of bias by capitalising on unique circumstances in which information about how studies were planned or conducted can be compared to how the studies were reported. Research findings reported in dissertations[23,167], regulatory records[168], ethics board approvals[169,170], third-party data collection records[171,172], and trial registries[173,174] were often more complete, smaller in magnitude, and less likely to be statistically significant, compared with results from the same studies reported in the published literature. Though the inherent potential of scientific methods to generate useful knowledge is not in doubt, evidence of the routine infusion of bias into the scientific ecosystem has prompted a multidisciplinary crisis of confidence[11–14].

Preregistration to reduce bias, increase transparency, and calibrate confidence

Preregistration involves declaring research plans in a public registry before research outcomes are known. Preregistration (1) reduces the risk of bias by encouraging outcome-independent decision-making; and (2) increases transparency, enabling others to assess the risk of bias and calibrate their confidence in research outcomes. To clarify these two core functions, it is helpful to appreciate that the extent of prespecification

provided by a preregistration can vary along a spectrum (Figure 2). A minimal preregistration may prespecify very few research decisions, thus providing little constraint on outcome-dependent decision-making. This entails a higher risk of bias, but also increases sensitivity to serendipitous discovery (all else being equal). Minimal preregistrations are therefore entirely appropriate in more *exploratory* research contexts where researchers are mostly concerned with generating hypotheses. By contrast, a maximal preregistration may prespecify all relevant research decisions, thus providing strong constraint on outcome-dependent decision-making. This entails a lower risk of bias, but also decreases sensitivity to serendipitous discovery (all else being equal). Maximal preregistrations are therefore most appropriate in more *confirmatory* research contexts where researchers are mostly concerned with testing hypotheses. In practice, research activities may fall at any point along this spectrum and studies may contain both exploratory and confirmatory elements[16,175].

Crucially, regardless of the degree of prespecification, preregistration is valuable because it transparently provides the context readers need to calibrate their confidence in research outcomes. By making clear which aspects of the research were outcome-independent (i.e., more confirmatory) and which were outcome-dependent (i.e., more exploratory), preregistration helps researchers understand uncertainty in the outcomes arising from risk of bias[16,176]. Specifically, all else being equal, researchers should be skeptical of outcomes derived from exploratory research because they have a higher risk of bias[177,178]. Our confidence in research should of course be influenced by other factors as well (e.g., the validity of the statistical methods, appropriate implementation of randomization, theoretical rationale, etc.), but this is beyond the scope of preregistration. Having the necessary information to appreciate the risk of bias is important for all research consumers (including the original researchers) and should be incorporated into formal evidence synthesis (meta-analyses and systematic reviews)[177,179]. It is important, for example, that practitioners know how much confidence to have in a research claim when deciding whether to deploy an evidence-based policy intervention[9] or administer a medical treatment[10].

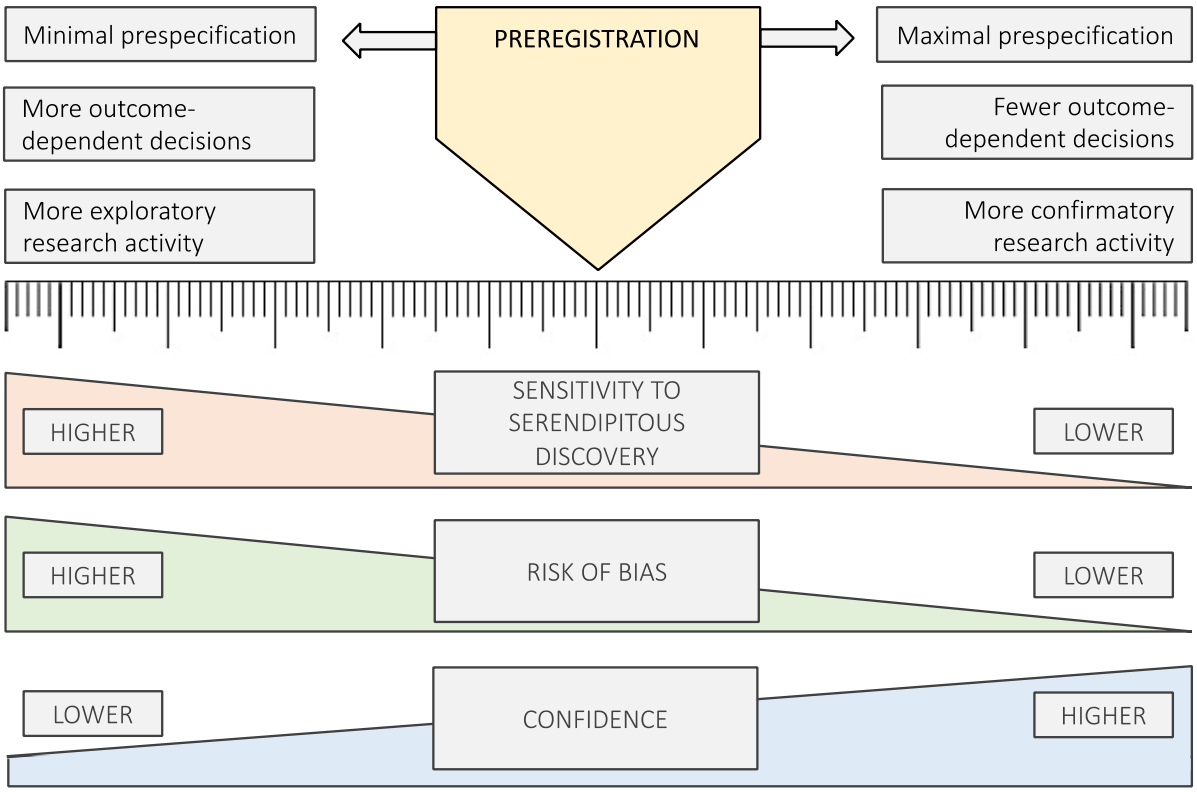

Figure 2. Schematic to illustrate that the degree of prespecification provided by a preregistration can vary along a minimal-maximal continuum depending on the extent to which researcher degrees of freedom are constrained. All else being equal, the degree of prespecification impacts sensitivity to serendipitous discovery and the risk of bias inherent to a particular research activity. Individual studies may contain both exploratory and confirmatory research activities. Preregistration transparently communicates where particular research outcomes are located along the continuum, helping readers to appropriately calibrate their confidence.

Box 4. Meta-research: Preregistration in practice

Even good policy ideas, promoted with the best of intentions, can be less effective than hoped for, or even have unintended negative consequences[180]. Whether the theoretical benefits of preregistration are realised in practice will depend on how it is implemented. It is therefore prudent to proceed with caution and conduct meta-research to empirically evaluate and monitor preregistration in practice[181]. Below we highlight selected insights based on existing meta-research on preregistration. Note that a general limitation of these studies is the use of observational designs in idiosyncratic contexts, limiting generalisability and precluding straightforward causal conclusions.

- Mandates to register clinical trials have been associated with a substantial increase in trial registration[182,183], though some trials are still registered retrospectively[48,184].
- Preregistrations often lack detail or are ambiguously specified in medicine[48–50], economics[70,185], and on the Open Science Framework[95].
- Unacknowledged discrepancies between preregistrations and published reports appear to be common in several domains including clinical trials[46,47], psychology[186], and economics[70] (for review see ref[187]).
- Preregistered clinical trials have been associated with lower risk of bias evaluations[188].
- Several observational studies in medicine[189–191] and psychology[40,192,193] have found a lower proportion of 'statistically significant' findings and smaller effects reported in preregistered research relative to non-preregistered research; however, this association has not been observed in at least one other study[194].

Some researchers have voiced concern that preregistration devalues or even prevents exploratory research[64,65,70,89,198]; however, this is not intended. Exploratory (outcome-dependent) and confirmatory (outcome-independent) research activities have complementary strengths and weaknesses, and both make important epistemic contributions[199,200] (a related distinction in philosophy is the 'context of discovery' and the 'context of justification')[201]. Exploratory research activities are most suitable for generating hypotheses because they are more sensitive to serendipitous discovery. However, this virtue is also their Achilles' heel because it incurs a higher risk of bias. Confirmatory research activities on the other hand, are most suitable for testing

hypotheses because they have a lower risk of bias; their weakness is a lack of sensitivity to serendipitous discovery. Fortunately, preregistration affords researchers the best of both worlds. Within an individual study, researcher degrees of freedom can be constrained for the purposes of confirmatory analyses, and relaxed for the purposes of exploratory analyses. Preregistration enables differentiation between the two so confidence in the outcomes of each type of research can be appropriately calibrated. Clarifying the exploratory-confirmatory distinction may also reduce the pressure on researchers to disguise exploratory research as confirmatory research[23,126], potentially exposing that many extant theories and measurement tools are not sufficiently developed to enable informative hypothesis tests[89,107,202]. In sum, both exploratory and confirmatory research are valuable; preregistration enables us to determine which is which.

In addition to the two core functions (reducing bias and increasing transparency) described above, preregistration may have several auxiliary benefits which include encouraging closer attention to study planning[193,195]; creating opportunity for peer feedback before study commencement (particularly with Registered Reports, Box 2); improved communication between collaborators[195,196]; increased discoverability of in-progress studies (potentially aiding participant recruitment, reducing waste from duplicated research, and enhancing efficiency through shared resources and collaboration)[7]; and protection against pressure from industry partners[51,61,62], reviewers[197], or collaborators[155] to obtain different research outcomes.

Recommendations

Preregistration should not be treated as a panacea

Preregistration is one tool in a scientist's toolbox, with specific pragmatic functions to reduce bias, increase transparency, and calibrate confidence. It does not guarantee study quality or sound inference, nor does it directly improve poorly specified theories[57,89,107]. Complementary tools can be used together to maximise their joint impact (Box 2). For example, embedding preregistration within the publication pipeline ('Registered Reports') may provide enhanced protection against publication bias[76]. Nevertheless, tools like preregistration will likely not be enough to address the multitude of issues that

undermine research utility and credibility; cultural and institutional changes are needed to address academia's skewed incentive structures[181,206] (Box 1).

Preregistrations should be as comprehensive as possible and as flexible as necessary

To take advantage of preregistration's bias-reduction function, researchers should aim to prespecify as many research decisions as they are capable of doing so in line with their exploratory or confirmatory goals. Preregistration templates and reporting guidelines may help researchers identify the most relevant decisions they need to make for a particular research design. However, in practice, researchers may find it difficult to prespecify all research decisions, especially in more exploratory contexts that inherently involve outcome-dependent decision-making. Additionally, researchers may need to depart from prespecified plans in order to handle unanticipated events[56]. This is all perfectly acceptable; one only needs to recognise that more comprehensive preregistrations provide greater protection from bias and engaging in outcome-dependent decision-making will increase the risk of bias. Even if researchers have not prespecified any research decisions, a preregistration that explicitly states the absence of a plan remains useful because it transparently communicates the exploratory nature of the research, helping others to evaluate the risk of bias and calibrate their confidence in the research outcomes (Figure 2). Additionally, prespecifying that a research activity is exploratory protects against that research being subsequently reframed as confirmatory[23,126], which could give readers false confidence in the outcomes.

Departures from the preregistered plan should be tolerated and evaluated

Departures from the preregistered plan can be acceptable or even desirable[89], as long as they are transparently recorded (e.g., in an open lab notebook[99], Box 2) and highlighted in published articles (e.g., https://osf.io/xv5rp/). For example, if a researcher realises during or after completion of the study that the preregistered analysis plan is ill-conceived, then it makes sense to depart from the plan and switch to a more appropriate method of analysis. It is crucial to recognise that if departures from the plan are made in an outcome-dependent manner, then this can increase the risk of bias, regardless of whether it yields other benefits. There are several tools that can be used to simultaneously benefit from plan departures whilst reducing the risk of bias (Box 2).

Preregistered blind analysis can be used preemptively to allow for handling unanticipated events without engaging in outcome-dependent decision making. If the researcher has already observed research outcomes, blind analysis could be requested from an independent 3rd party who has not observed the data. Another option is to perform robustness checks to evaluate whether the results are substantially affected by the use of different analysis procedures. Alternatively, it may be necessary to run a confirmatory analysis with a fresh dataset. These may also be useful strategies when researchers intend to conduct secondary analyses with datasets from which they are already aware of research outcomes, and thus cannot make outcome-independent decisions.

Preregistrations will require nuanced interpretation

Preregistration is intended to facilitate scientific judgement, not replace it. Evaluating preregistrations and understanding their impact on research outcomes requires time and expertise. It is not sufficient to know that "the study was preregistered"; one needs to know exactly which aspects of the study were preregistered, whether there were departures from the plan, and the impact on risk of bias. Interpretation will often require domain-specific expertise and evaluating preregistrations may be a skill that takes time to develop[56].

Journals, editors, and reviewers will have a key role in encouraging adoption and providing quality control

Journals could have a key role in encouraging adoption of preregistration. Major medical journals[43] have required registration of clinical trials since 2005. By contrast, a recent analysis of Transparency and Openness Promotion (TOP) Guidelines adoption suggests that fewer than 10% of psychology journals have any kind of preregistration policy[71]. As preregistration is rare beyond clinical trials[74,75], some caution, community consultation, and policy experimentation[181] may be desirable before broader preregistration mandates are considered. Ideally, preregistration will become a cultural norm because researchers believe it is a useful tool. Nevertheless, journal policies may serve to explicitly signal community norms. At this stage, an incipient journal policy (TOP Level 1) might simply require a transparency statement in which authors declare whether study plans were preregistered or not[205].

The effectiveness of preregistration will depend on its implementation, highlighting the importance of quality control. Most research consumers will probably not read study preregistrations. Indeed, some research suggests that clinical trial registry information is often not even read by peer reviewers[203]. Minimally, it would be helpful for journals to clarify their expectations regarding reviewer evaluation of preregistrations and reviewers could be asked to confirm whether or not they have examined preregistrations. Editors and reviewers could also receive guidance on how to evaluate preregistrations, particularly how to appraise departures from the original plan in terms of their impact on risk of bias and strategies to reduce risk of bias even if departures have occurred. Development and empirical evaluation of tools to support dedicated preregistration review is underway[204]. To enable some degree of quality control whilst conserving resources, journals could consider randomly allocating articles to undergo preregistration review. Identifying effective and efficient ways to implement quality control of preregistrations is an important topic for future meta-research.

Maximise efficiency by aligning tools for the planning, reporting, and evaluation of research

Preregistration can increase administrative burden because of the additional information that needs to be documented and evaluated[195,207]. The structure and content of preregistration templates, manuscript formats, reporting guidelines, and risk of bias tools should be aligned as much as possible to maximise the efficiency of writing, reporting, and evaluating preregistrations. For example, the recently introduced PRP-QUANT template[196] (Supplementary Information A) is aligned with APA manuscript formatting requirements and APA-JARS reporting guidelines[208]. This will facilitate transfer and cross referencing of information between the preregistration, manuscript, and reporting guidelines, thereby reducing redundancy and aiding evaluation. Embedding meta-data in these documents will also help to improve cross referencing and potentially enable some degree of automated validation (e.g., detection of unacknowledged plan departures). Further efforts towards tool alignment could learn from the EQUATOR model for the development and organisation of reporting guidelines in health research[209]. Templates and guidelines could be created for specific research designs based on consultation with the relevant research communities (using for

example, Delphi methods[210]). Risk of bias tools developed in medicine[177,179] could also be adapted for use in non-medical domains, helping to enhance and standardise the evaluation of preregistrations during peer review and formal evidence synthesis (systematic reviews and meta-analysis).

Evaluate and monitor preregistration with meta-research

Continual evaluation of how preregistration is implemented in various contexts will be vital to monitor its effectiveness and remain responsive to any unintended negative side-effects[180,181]. Policy makers (e.g., journal editors) could partner with meta-researchers to formally evaluate different policy options. Key questions may include: does preregistration hamper exploratory research[64,65,70,89,198], create burdensome bureaucracy[207], or give research "a superficial veneer of rigor"[58]? Though preregistration proponents have offered conceptual rebuttals to concerns like these[17,60,108,211], they will all depend on how preregistration is implemented in practice, which can be assessed empirically. A scoping review may be an important initial step to collate extant evidence and identify knowledge gaps. It will likely be productive to conduct randomized trials to evaluate the effectiveness of preregistration and other complementary tools where possible (Box 2). Key questions may include: to what extent does preregistration reduce bias? Do research consumers use preregistration to calibrate their confidence? Are study outcomes less biased and more representative when preregistration is combined with robustness checks?

Conclusion

Scientific methods are humanity's most effective means for learning about the world; however, science does not guarantee truth. Research is performed by fallible humans and every study has the potential to mislead as well as enlighten. Success, in part, depends on reducing bias and being transparent so we can more effectively calibrate our confidence in scientific claims. Preregistration is a pragmatic tool that has the potential to facilitate both of these goals and contribute to a "long history of learning how to not fool ourselves"[15].

Funding statement

Tom Hardwicke receives funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 841188.

Conflict of interest statement

The authors declare no conflicts of interest.

References

1. Bacon, F. *Novum organum*. (1620).
2. Gilovich, T. *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life*. (Free Press, 1991).
3. Laplace, P. S. *Essai Philosophique sur les Probabilités*. (Springer, 1825).
4. Nosek, B. A., Spies, J. R. & Motyl, M. Scientific Utopia: II. restructuring incentives and practices to promote truth over publishability. *Perspect. Psychol. Sci.* **7**, 615–631 (2012).
5. Smaldino, P. E. & McElreath, R. The natural selection of bad science. *R. Soc. Open Sci.* **3**, 160384 (2016).
6. Young, N. S., Ioannidis, J. P. A. & Al-Ubaydli, O. Why current publication practices may distort science. *PLOS Med.* **5**, e201 (2008).
7. Glasziou, P. *et al.* Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet* **383**, 267–276 (2014).
8. Nissen, S. B., Magidson, T., Gross, K. & Bergstrom, C. T. Publication bias and the canonization of false facts. *eLife* **5**, e21451 (2016).
9. IJzerman, H. *et al.* Use caution when applying behavioural science to policy. *Nat. Hum. Behav.* **4**, 1092–1094 (2020).
10. Prasad, V. K. & Cifu, A. S. *Ending Medical Reversal: Improving Outcomes, Saving Lives*. (John Hopkins University Press, 2019).

11. Christensen, G. & Miguel, E. Transparency, reproducibility, and the credibility of economics research. *J. Econ. Lit.* **56**, 920–980 (2018).

12. Pashler, H. & Wagenmakers, E.-J. Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? *Perspect. Psychol. Sci.* **7**, 528–530 (2012).

13. Baker, M. 1,500 scientists lift the lid on reproducibility. *Nat. News* **533**, 452 (2016).

14. Ioannidis, J. P. A. Why most published research findings are false. *PLOS Med.* **2**, e124 (2005).

15. Feynman, R. P. Cargo Cult Science. (1974).

16. Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J. & Kievit, R. A. An agenda for purely confirmatory research. *Perspect. Psychol. Sci.* **7**, 632–638 (2012).

17. Nosek, B. A., Ebersole, C. R., DeHaven, A. C. & Mellor, D. T. The preregistration revolution. *Proc. Natl. Acad. Sci.* **115**, 2600–2606 (2018).

18. Veldkamp, C. L. S., Hartgerink, C. H. J., Assen, M. A. L. M. van & Wicherts, J. M. Who believes in the storybook image of the scientist? *Account. Res.* **24**, 127–151 (2017).

19. Barber, T. X. *Pitfalls in Human Research: Ten Pivotal Points*. (Pergamon Press, 1976).

20. Bakker, M., van Dijk, A. & Wicherts, J. M. The rules of the game called psychological science. *Perspect. Psychol. Sci.* **7**, 543–554 (2012).

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

22. Giner-Sorolla, R. Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspect. Psychol. Sci.* **7**, 562–571 (2012).

23. O'Boyle, E. H., Banks, G. C. & Gonzalez-Mulé, E. The chrysalis effect: how ugly initial results metamorphosize into beautiful articles. *J. Manag.* **43**, 376–399 (2017).

24. Brewer, W. F. & Chinn, C. A. Scientists' responses to anomalous data: evidence from psychology, history, and philosophy of science. *PSA Proc. Bienn. Meet. Philos. Sci. Assoc.* **1994**, 304–313 (1994).

25. Edwards, K. & Smith, E. E. A disconfirmation bias in the evaluation of arguments. *Attitudes Soc. Cogn.* **71**, 5–24 (1996).

26. Mynatt, C. R., Doherty, M. E. & Tweney, R. D. Confirmation bias in a simulated research environment: an experimental study of scientific inference: *Q. J. Exp. Psychol.* (1977) doi:10.1080/00335557743000053.

27. Hoekstra, R. & Vazire, S. Aspiring to greater intellectual humility in science. *Nat. Hum. Behav.* 1–6 (2021) doi:10.1038/s41562-021-01203-8.

28. Nickerson, R. S. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* **2**, 175–220 (1998).

29. Kunda, Z. The case for motivated reasoning. *Psychol. Bull.* **108**, 480–498 (1990).

30. Fischhoff, B. Hindsight does not equal foresight: the effect of outcome knowledge on judgment under uncertainty. *J. Exp. Psychol. Hum. Percept. Perform.* **1**, 288–299 (1975).

31. Slovic, P. & Fischhoff, B. On the psychology of experimental surprises. *J. Exp. Psychol. Hum. Percept. Perform.* **3**, 544–551 (1977).

32. Gilovich, T., Vallone, R. & Tversky, A. The hot hand in basketball: On the misperception of random sequences. *Cognit. Psychol.* **17**, 295–314 (1985).

33. Pronin, E. Perception and misperception of bias in human judgment. *Trends Cogn. Sci.* **11**, 37–43 (2007).

34. de Groot, A. D. The meaning of "significance" for different types of research. *Acta Psychol. (Amst.)* **148**, 188–194 (2014).

35. Keynes, J. M. *Treatise on Probability*. (Macmillan & Co, 1921).

36. Mill, J. S. *A system of logic*. (George Routledge & Sons, 1843).

37. Peirce, C. S. A theory of probable inference. in *Studies in logic* (ed. Peirce, C. S.) 126–281 (Little, Brown, 1883).

38. Bakan, D. The test of significance in psychological research. *Psychol. Bull.* **66**, 423–437 (1966).

39. Walster, G. W. & Cleary, T. A. A proposal for a new editorial policy in the social sciences. *Am. Stat.* **24**, 16–19 (1970).

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

41. Spellman, B. A. A short (personal) future history of Revolution 2.0. *Perspect. Psychol. Sci.* **10**, 886–899 (2015).

42. Simes, R. J. Publication bias: the case for an international registry of clinical trials. *J. Clin. Oncol.* **4**, 1529–1541 (1986).

43. De Angelis, C. *et al.* Clinical trial registration: A statement from the International Committee of Medical Journal Editors. *N. Engl. J. Med.* **351**, 1250–1251 (2004).

44. Dickersin, K. & Rennie, D. The evolution of trial registries and their use to assess the clinical trial enterprise. *JAMA* **307**, 1861–1864 (2012).

45. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* **310**, 2191 (2013).

46. Chan, A.-W., Hróbjartsson, A., Haahr, M. T., Gøtzsche, P. C. & Altman, D. G. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* **291**, 2457–2465 (2004).

47. Dwan, K. *et al.* Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. *PLOS Med.* **11**, e1001666 (2014).

48. Zarin, D. A., Tse, T., Williams, R. J. & Rajakannan, T. Update on trial registration 11 years after the ICMJE policy was established. *N. Engl. J. Med.* **376**, 383–391 (2017).

49. Greenberg, L., Jairath, V., Pearse, R. & Kahan, B. C. Pre-specification of statistical analysis approaches in published clinical trial protocols was inadequate. *J. Clin. Epidemiol.* **101**, 53–60 (2018).

50. Chan, A.-W., Hróbjartsson, A., Jørgensen, K. J., Gøtzsche, P. C. & Altman, D. G. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ* **337**, a2299 (2008).

51. Humphreys, M., Sierra, R. S. de la & Windt, P. van der. Fishing, commitment, and communication: a proposal for comprehensive nonbinding research registration. *Polit. Anal.* **21**, 1–20 (2013).

52. Dal-Ré, R. *et al.* Making prospective registration of observational research a reality. *Sci. Transl. Med.* **6**, 1–4 (2014).

53. Lash, T. L. Preregistration of Study Protocols Is Unlikely to Improve the Yield From Our Science, But Other Strategies Might. *Epidemiology* **21**, 612–613 (2010).

54. Boccia, S. *et al.* Registration practices for observational studies on ClinicalTrials.gov indicated low adherence. *J. Clin. Epidemiol.* **70**, 176–182 (2016).

55. Weber, W. E. J., Merino, J. G. & Loder, E. Trial registration 10 years on. *BMJ* **351**, (2015).

56. Nosek, B. A. *et al.* Preregistration is hard, and worthwhile. *Trends Cogn. Sci.* **23**, 815–818 (2019).

57. Szollosi, A. *et al.* Is preregistration worthwhile? *Trends Cogn. Sci.* **24**, 94–95 (2020).

58. Devezer, B., Navarro, D. J., Vandekerckhove, J. & Buzbas, E. O. The case for formal methodology in scientific reform. *R. Soc. Open Sci.* **8**, 200805 (2020).

59. Rubin, M. Does preregistration improve the credibility of research findings? *Quant. Methods Psychol.* **16**, 15 (2020).

60. Simmons, J. P., Nelson, L. D. & Simonsohn, U. Pre-registration: why and how. *J. Consum. Psychol.* **31**, 151–162 (2021).

61. Monogan, J. E. Research preregistration in political science: the case, counterarguments, and a response to critiques. *PS Polit. Sci. Polit.* **48**, 425–429 (2015).

62. Olken, B. A. Promises and perils of pre-analysis plans. *J. Econ. Perspect.* **29**, 61–80 (2015).

63. Casey, K., Glennerster, R. & Miguel, E. Reshaping institutions: evidence on aid impacts using a preanalysis plan*. *Q. J. Econ.* **127**, 1755–1812 (2012).

64. Coffman, L. C. & Niederle, M. Pre-analysis plans have limited upside, especially where replications are feasible. *J. Econ. Perspect.* **29**, 81–98 (2015).

65. Shiffrin, R. M. Commentary on "robust modeling in cognitive science: misunderstanding the goal of modeling". *Comput. Brain Behav.* **2**, 176–178 (2019).

66. Crüwell, S. & Evans, N. J. Preregistration in diverse contexts: a preregistration template for the application of cognitive models. *R. Soc. Open Sci.* **8**, 210155 (2021).

67. Paul, M., Govaart, G. H. & Schettino, A. Making ERP research more transparent: guidelines for preregistration. *Int. J. Psychophysiol.* (2021) doi:10.1016/j.ijpsycho.2021.02.016.

68. Akker, O. van den *et al. Preregistration of secondary data analysis: A template and tutorial*. https://psyarxiv.com/hvfmr/ (2019).

69. Haven, T. L. & Grootel, D. L. V. Preregistering qualitative research. *Account. Res.* **26**, 229–244 (2019).

70. Ofosu, G. K. & Posner, D. N. Pre-analysis plans: an early stocktaking. *Perspect. Polit.* 1–17 (2020) doi:10.1017/S1537592721000931.

71. Nosek, B. A. *et al.* Replicability, robustness, and reproducibility in psychological science. *Annu. Rev. Psychol.* **73**, annurev-psych-020821-114157 (2022).

72. Christensen, G. *et al.* Open science practices are on the rise: the state of social science (3s) survey. (2019) doi:10.31222/osf.io/5rksu.

73. Wallach, J. D., Boyack, K. W. & Ioannidis, J. P. A. Reproducible research practices, transparency, and open access data in the biomedical literature, 2015–2017. *PLOS Biol.* **16**, e2006930 (2018).

74. Hardwicke, T. E. *et al.* Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014-2017). *Perspect. Psychol. Sci.* (2021) doi:https://doi.org/10.1177/1745691620979806.

75. Hardwicke, T. E. *et al.* An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014–2017). *R. Soc. Open Sci.* **7**, 190806 (2020).

76. Chambers, C. D. & Tzavella, L. The past, present and future of Registered Reports. *Nat. Hum. Behav.* 1–14 (2021) doi:10.1038/s41562-021-01193-7.

77. Montoya, A. K., Krenzer, W. L. D. & Fossum, J. L. Opening the door to registered reports: census of journals publishing registered reports (2013–2020). *Collabra Psychol.* **7**, (2021).

78. Hardwicke, T. E. & Ioannidis, J. P. A. Mapping the universe of registered reports. *Nat. Hum. Behav.* **2**, 793–796 (2018).

79. Srivastava, S. Sound inference in complicated research: a multi-strategy approach. (2018) doi:10.31234/osf.io/bwr48.

80. Baldwin, J. R., Pingault, J.-B., Schoeler, T., Sallis, H. M. & Munafò, M. R. Protecting against researcher bias in secondary data analysis: challenges and potential solutions. *Eur. J. Epidemiol.* 1–10 (2022) doi:10.1007/s10654-021-00839-0.

81. Thabane, L. *et al.* A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med. Res. Methodol.* **13**, 92 (2013).

82. Steegen, S., Tuerlinckx, F., Gelman, A. & Vanpaemel, W. Increasing transparency through a multiverse analysis. *Perspect. Psychol. Sci.* **11**, 702–712 (2016).

83. Patel, C. J., Burford, B. & Ioannidis, J. P. A. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *J. Clin. Epidemiol.* **68**, 1046–1058 (2015).

84. Simonsohn, U., Simmons, J. P. & Nelson, L. D. Specification curve analysis. *Nat. Hum. Behav.* 1–7 (2020) doi:10.1038/s41562-020-0912-z.

85. Young, C. Model uncertainty and the crisis in science. *Socius* **4**, 1–7 (2018).

86. Haaf, J. M., Hoogeveen, S., Berkhout, S., Gronau, Q. F. & Wagenmakers, E.-J. A bayesian multiverse analysis of Many Labs 4: quantifying the evidence against mortality salience. (2020) doi:10.31234/osf.io/cb9er.

87. Klau, S., Hoffmann, S., Patel, C. J., Ioannidis, J. P. A. & Boulesteix, A.-L. Examining the robustness of observational associations to model, measurement and sampling uncertainty with the vibration of effects framework. *Int. J. Epidemiol.* **50**, 266–278 (2021).

88. Orben, A. & Przybylski, A. K. The association between adolescent well-being and digital technology use. *Nat. Hum. Behav.* **3**, 173–182 (2019).

89. Oberauer, K. & Lewandowsky, S. Addressing the theory crisis in psychology. *Psychon. Bull. Rev.* **26**, 1596–1618 (2019).

90. Giudice, M. D. & Gangestad, S. A traveler's guide to the multiverse: promises, pitfalls, and a framework for the evaluation of analytic decisions. *Adv. Methods Pract. Psychol. Sci.* **4**, 1–15 (2021).

91. Dutilh, G., Sarafoglou, A. & Wagenmakers, E.-J. Flexible yet fair: blinding analyses in experimental psychology. *Synthese* (2019) doi:https://doi.org/10.1007/s11229-019-02456-7.

92. MacCoun, R. & Perlmutter, S. Blind analysis: Hide results to seek the truth. *Nat. News* **526**, 187 (2015).

93. Klein, J. R. & Roodman, A. Blind analysis in nuclear and particle physics. *Annu. Rev. Nucl. Part. Sci.* **55**, 141–163 (2005).

94. Schorfheide, F. & Wolpin, K. I. On the use of holdout samples for model selection. *Am. Econ. Rev.* **102**, 477–481 (2012).

95. Bakker, M. *et al.* Ensuring the quality and specificity of preregistrations. *PLOS Biol.* **18**, e3000937 (2020).

96. Hardwicke, T. E. *et al.* Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal Cognition. *R. Soc. Open Sci.* **5**, 180448 (2018).

97. Hardwicke, T. E. *et al.* Analytic reproducibility in articles receiving open data badges at the journal Psychological Science: an observational study. *R. Soc. Open Sci.* **8**, 201494 (2021).

98. Lin, W. & Green, D. P. Standard operating procedures: a safety net for pre-analysis plans. *PS Polit. Sci. Polit.* **49**, 495–500 (2016).

99. Navarro, D. *Paths in strange spaces: A comment on preregistration*. https://psyarxiv.com/wxn58/ (2020).

100. McIntosh, R. D. Exploratory reports: A new article type for Cortex. *Cortex* **96**, A1–A4 (2017).

101. Button, K. S., Bal, L., Clark, A. & Shipley, T. Preventing the ends from justifying the means: withholding results to address publication bias in peer-review. *BMC Psychol.* **4**, 1–7 (2016).

102. DeVito, N. J., Bacon, S. & Goldacre, B. Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: a cohort study. *The Lancet* **395**, 361–369 (2020).

103. Wieschowski, S. *et al.* Result dissemination from clinical trials conducted at German university medical centers was delayed and incomplete. *J. Clin. Epidemiol.* **115**, 37–45 (2019).

104. Leamer, E. E. Let's take the con out of econometrics. *Am. Econ. Rev.* **73**, 31–43 (1983).

105. Page, M. J., McKenzie, J. E. & Forbes, A. Many scenarios exist for selective inclusion and reporting of results in randomized trials and systematic reviews. *J. Clin. Epidemiol.* **66**, 524–537 (2013).

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

107. Szollosi, A. & Donkin, C. Arrested theory development: the misguided distinction between exploratory and confirmatory research. *Perspect. Psychol. Sci.* 1745691620966796 (2021) doi:10.1177/1745691620966796.

108. Forstmeier, W., Wagenmakers, E.-J. & Parker, T. H. Detecting and avoiding likely false-positive findings – a practical guide. *Biol. Rev.* **92**, 1941–1968 (2017).

109. Gelman, A. & Loken, E. The statistical crisis in science. *Am. Sci.* **102**, 460–465 (2014).

110. Duhem, P. *The Aim and Structure of Physical Theory*. (Princeton University Press, 1954).

111. Gigerenzer, G. Surrogates for theories. *Theory Psychol.* **8**, 195–204 (1998).

112. Meehl, P. E. Theory-testing in psychology and physics: a methodological paradox. *Philos. Sci.* **34**, 103–115 (1967).

113. Muthukrishna, M. & Henrich, J. A problem in theory. *Nat. Hum. Behav.* **3**, 221–229 (2019).

114. Sackett, D. L. Bias in analytic research. *J. Chronic Dis.* **32**, 51–63 (1979).

115. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. (2011).

116. DeCoster, J., Sparks, E. A., Sparks, J. C., Sparks, G. G. & Sparks, C. W. Opportunistic biases: Their origins, effects, and an integrated solution. *Am. Psychol.* **70**, 499–514 (2015).

117. André, Q. Outlier exclusion procedures must be blind to the researcher's hypothesis. *J. Exp. Psychol. Gen.* No Pagination Specified-No Pagination Specified (2021) doi:10.1037/xge0001069.

118. Goldacre, B. *et al.* COMPare: a prospective cohort study correcting and monitoring 58 misreported trials in real time. *Trials* **20**, 118 (2019).

119. Cramer, A. O. J. *et al.* Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychon. Bull. Rev.* **23**, 640–647 (2016).

120. White, T. K., Reiter, J. P. & Petrin, A. Imputation in U.S. manufacturing data and its implications for productivity dispersion. *Rev. Econ. Stat.* **100**, 502–509 (2018).

121. Wallach, J. D. *et al.* Evaluation of evidence of statistical support and corroboration of subgroup claims in randomized clinical trials. *JAMA Intern. Med.* **177**, 554 (2017).

122. Armitage, P., McPherson, C. K. & Rowe, B. C. Repeated significance tests on accumulating data. *J. R. Stat. Soc. Ser. Gen.* **132**, 235–244 (1969).

123. Page, M. J. & Higgins, J. P. T. Rethinking the assessment of risk of bias due to selective reporting: a cross-sectional study. *Syst. Rev.* **5**, 108 (2016).

124. Babbage, C. *Reflections on the decline of science in England, and on some of its causes*. (Franklin Classics, 1830).

125. Goldacre, B. *Bad science*. (Fourth Estate, 2008).

126. Kerr, N. L. HARKing: Hypothesizing After the Results are Known. *Personal. Soc. Psychol. Rev. Lawrence Erlbaum Assoc.* **2**, 196 (1998).

127. Rubin, M. When does HARKing hurt? Identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. *Rev. Gen. Psychol.* **21**, 308–320 (2017).

128. Barnes, E. C. *The Paradox of Predictivism*. (Cambridge University Press, 2008).

129. Douglas, H. & Magnus, P. D. State of the Field: Why novel prediction matters. *Stud. Hist. Philos. Sci. Part A* **44**, 580–589 (2013).

130. Howson, C. Fitting your theory to the facts: probably not such a bad thing after all. in *Scientific Theories* (ed. Savage, C. W.) 224–244 (University of Minnesota Press, 1990).

131. Hitchcock, C. & Sober, E. Prediction versus accommodation and the risk of overfitting. *Br. J. Philos. Sci.* **55**, 1–34 (2004).

132. Lipton, P. Testing hypotheses: prediction and prejudice. *Science* **307**, 219–221 (2005).

133. Lakens, D. The value of preregistration for psychological science: a conceptual analysis. *Jpn. Psychol. Rev.* **62**, 272–280 (2019).

134. Mayo, D. G. *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. (Cambridge University Press, 2018).

135. Vanpaemel, W. The really risky registered modeling report: incentivizing strong tests and HONEST modeling in cognitive science. *Comput. Brain Behav.* **2**, 218–222 (2019).

136. Maher, P. Prediction, accommodation, and the logic of discovery. *PSA Proc. Bienn. Meet. Philos. Sci. Assoc.* **1988**, 273–285 (1988).

137. Worrall, J. Prediction and accommodation revisited. *Stud. Hist. Philos. Sci. Part A* **45**, 54–61 (2014).

138. Popper, K. *Conjectures and Refutations: The Growth of Scientific Knowledge*. (Harper & Row, 1963).

139. Lakatos, I. Falsification and the methodology of scientific research programmes. in *Criticism and the Growth of Knowledge* (eds. Musgrave, A. & Lakatos, I.) vol. 4 91–196 (Cambridge University Press, 1970).

140. Greenland, S. Analysis goals, error-cost sensitivity, and analysis hacking: Essential considerations in hypothesis testing and multiple comparisons. *Paediatr. Perinat. Epidemiol.* **35**, 8–23 (2021).

141. Scott, J. G. & Berger, J. O. Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Ann. Stat.* **38**, 2587–2619 (2010).

142. Dienes, Z. Bayesian versus orthodox statistics: which side are you on? *Perspect. Psychol. Sci.* **6**, 274–290 (2011).

143. Howson, C. & Urbach, P. *Scientific Reasoning: The Bayesian Approach*. (Open Court, 2006).

144. Dienes, Z. How Bayes factors change scientific practice. *J. Math. Psychol.* **72**, 78–89 (2016).

145. Wasserstein, R. L. & Lazar, N. A. The ASA's statement on p-values: context, process, and purpose. *Am. Stat.* **70**, 129–133 (2016).

146. Stefan, A. & Schönbrodt, F. Big little lies: a compendium and simulation of p-hacking strategies. (2022) doi:10.31234/osf.io/xy2dk.

147. Austin, P. C., Mamdani, M. M., Juurlink, D. N. & Hux, J. E. Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. *J. Clin. Epidemiol.* **59**, 964–969 (2006).

148. Bennett, C., Miller, M. & Wolford, G. Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: an argument for multiple comparisons correction. *NeuroImage* **47**, S125 (2009).

149. Botvinik-Nezer, R. *et al.* Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* **582**, 84–88 (2020).

150. Silberzahn, R. *et al.* Many analysts, one data set: making transparent how variations in analytic choices affect results: *Adv. Methods Pract. Psychol. Sci.* (2018) doi:10.1177/2515245917747646.

151. van Dongen, N. N. N. *et al.* Multiple perspectives on inference for two simple statistical scenarios. *Am. Stat.* **73**, 328–339 (2019).

152. Fanelli, D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLOS ONE* **4**, e5738 (2009).

153. Fraser, H., Parker, T., Nakagawa, S., Barnett, A. & Fidler, F. Questionable research practices in ecology and evolution. *PLOS ONE* **13**, e0200303 (2018).

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

155. Wang, M. Q., Yan, A. F. & Katz, R. V. Researcher requests for inappropriate analysis and reporting: a U.S. survey of consulting biostatisticians. *Ann. Intern. Med.* **169**, 554–558 (2018).

156. Chavalarias, D., Wallach, J. D., Li, A. H. T. & Ioannidis, J. P. A. Evolution of Reporting *P* Values in the Biomedical Literature, 1990-2015. *JAMA* **315**, 1141 (2016).

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

158. Sterling, T. D. Publication decisions and their possible effects on inferences drawn from tests of significance--or vice versa. *J. Am. Stat. Assoc.* **54**, 30–34 (1959).

159. Button, K. S. *et al.* Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14**, 365–376 (2013).

160. Szucs, D. & Ioannidis, J. P. A. Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biol.* **15**, e2000797 (2017).

161. Camerer, C. F. *et al.* Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nat. Hum. Behav.* **2**, 637–644 (2018).

162. Camerer, C. F. *et al.* Evaluating replicability of laboratory experiments in economics. *Science* **351**, 1433–1436 (2016).

163. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science* **349**, aac416 (2015).

164. Williamson, P. R., Gamble, C., Altman, D. G. & Hutton, J. L. Outcome selection bias in meta-analysis. *Stat. Methods Med. Res.* (2016) doi:10.1191/0962280205sm415oa.

165. Carter, E. C., Schönbrodt, F. D., Gervais, W. M. & Hilgard, J. Correcting for bias in psychology: a comparison of meta-analytic methods. *Adv. Methods Pract. Psychol. Sci.* **2**, 115–144 (2019).

166. Sterne, J. A. C. *et al.* Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* **343**, d4002 (2011).

167. Mazzola, J. J. & Deuling, J. K. Forgetting what we learned as graduate students: HARKing and selective outcome reporting in I–O journal articles. *Ind. Organ. Psychol.* **6**, 279–284 (2013).

168. Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A. & Rosenthal, R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N. Engl. J. Med.* **358**, 252–260 (2008).

169. Cooper, H., DeNeve, K. & Charlton, K. Finding the missing science: the fate of studies submitted for review by a human subjects committee. *Psychol. Methods* **2**, 447–452 (1997).

170. Hahn, S., Williamson, P. R. & Hutton, J. L. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *J. Eval. Clin. Pract.* **8**, 353–359 (2002).

171. Franco, A., Malhotra, N. & Simonovits, G. Underreporting in psychology experiments: evidence from a study registry. *Soc. Psychol. Personal. Sci.* **7**, 8–12 (2016).

172. Franco, A., Malhotra, N. & Simonovits, G. Publication bias in the social sciences: Unlocking the file drawer. *Science* **345**, 1502–1505 (2014).

173. Becker, J. E., Krumholz, H. M., Ben-Josef, G. & Ross, J. S. Reporting of results in ClinicalTrials.gov and high-impact journals. *JAMA* **311**, 1063–1065 (2014).

174. Hartung, D. M. *et al.* Reporting discrepancies between the clinicaltrials.gov results database and peer-reviewed publications. *Ann. Intern. Med.* **160**, 477–483 (2014).

175. Fife, D. & Rodgers, J. L. Understanding the exploratory/confirmatory data analysis continuum: moving beyond the 'replication crisis'. (2019) doi:10.31234/osf.io/5vfq6.

176. Kimmelman, J., Mogil, J. S. & Dirnagl, U. Distinguishing between Exploratory and Confirmatory Preclinical Research Will Improve Translation. *PLOS Biol.* **12**, e1001863 (2014).

177. Sterne, J. A. C. *et al.* RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* **366**, l4898 (2019).

178. Turner, L., Boutron, I., Hróbjartsson, A., Altman, D. G. & Moher, D. The evolution of assessing bias in Cochrane systematic reviews of interventions: celebrating methodological contributions of the Cochrane Collaboration. *Syst. Rev.* **2**, 79 (2013).

179. Page, M. J., McKenzie, J. E. & Higgins, J. P. T. Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. *BMJ Open* **8**, e019703 (2018).

180. Ioannidis, J. P. A. Handling the fragile vase of scientific practices. *Addiction* **110**, 9–10 (2015).

181. Hardwicke, T. E. *et al.* Calibrating the scientific ecosystem through meta-research. *Annu. Rev. Stat. Its Appl.* **7**, 11–37 (2020).

182. Zou, C. X. *et al.* Registration, results reporting, and publication bias of clinical trials supporting FDA approval of neuropsychiatric drugs before and after FDAAA: a retrospective cohort study. *Trials* **19**, 581 (2018).

183. Phillips, A. T. *et al.* Association of the FDA Amendment Act with trial registration, publication, and outcome reporting. *Trials* **18**, 333 (2017).

184. Dal-Ré, R., Ross, J. S. & Marušić, A. Compliance with prospective trial registration guidance remained low in high-impact journals and has implications for primary end point reporting. *J. Clin. Epidemiol.* **75**, 100–107 (2016).

185. Abrams, E., Libgober, J. & List, J. A. *Research registries: taking stock and looking forward*. 1–66 https://uploads.strikinglycdn.com/files/840dd740-4a8d-4f09-8dbd-e6498f5661c2/January2021Version.pdf (2021).

186. Claesen, A., Gomes, S. L. B. T., Tuerlinckx, F. & Vanpaemel, W. Preregistration: comparing dream to reality. (2019) doi:10.31234/osf.io/d8wex.

187. TARG Meta-Research Group *et al.* Estimating the prevalence of discrepancies between study registrations and publications: A systematic review and meta-analyses. *medRxiv* 2021.07.07.21259868 (2021) doi:10.1101/2021.07.07.21259868.

188. Tan, A. C. *et al.* Prevalence of trial registration varies by study characteristics and risk of bias. *J. Clin. Epidemiol.* **113**, 64–74 (2019).

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

190. Papageorgiou, S. N., Xavier, G. M., Cobourne, M. T. & Eliades, T. Registered trials report less beneficial treatment effects than unregistered ones: a meta-epidemiological study in orthodontics. *J. Clin. Epidemiol.* **100**, 44–52 (2018).

191. Kaplan, R. M. & Irvin, V. L. Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLOS ONE* **10**, e0132382 (2015).

192. Scheel, A. M., Schijen, M. & Lakens, D. An excess of positive results: Comparing the standard Psychology literature with Registered Reports. *Advances in Methods and Practices in Psychological Science* vol. 4 1–12 (2020).

193. Toth, A. A. *et al.* Study preregistration: an evaluation of a method for transparent reporting. *J. Bus. Psychol.* (2020) doi:10.1007/s10869-020-09695-3.

194. Odutayo, A. *et al.* Association between trial registration and positive study findings: cross sectional study (Epidemiological Study of Randomized Trials—ESORT). *BMJ* **356**, j917 (2017).

195. Sarafoglou, A., Kovacs, M., Bakos, B. E., Wagenmakers, E.-J. & Aczel, B. Is preregistration worthwhile? A survey on personal experiences. (2021) doi:10.31234/osf.io/6k5gr.

196. Bosnjak, M. *et al.* A template for preregistration of quantitative research in psychology: report of the joint psychological societies preregistration task force. (2021) doi:10.31234/osf.io/d7m5r.

197. Nosek, B. A. & Lakens, D. Registered reports: A method to increase the credibility of published results. *Soc. Psychol.* **45**, 137 (2014).

198. MacEachern, S. N. & Zandt, T. V. Preregistration of modeling exercises may not be useful. *Comput. Brain Behav.* **2**, 179–182 (2019).

199. Tukey, J. W. We need both exploratory and confirmatory. *Am. Stat.* **34**, 23–25 (1980).

200. Wagenmakers, E.-J., Dutilh, G. & Sarafoglou, A. The creativity-verification cycle in psychological science: new methods to combat old idols: *Perspect. Psychol. Sci.* (2018) doi:10.1177/1745691618771357.

201. Reichenbach, H. *Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge*. (University of Chicago Press, 1938).

202. Scheel, A. M., Tiokhin, L., Isager, P. M. & Lakens, D. Why hypothesis testers should spend less time testing hypotheses. *Perspect. Psychol. Sci.* (2020) doi:https://doi.org/10.1177/1745691620966795.

203. Mathieu, S., Chan, A.-W. & Ravaud, P. Use of trial register information during the peer review process. *PLOS ONE* **8**, e59910 (2013).

204. Thibault, R. T., Drax, K., Thompson, J. & Munafo, M. A peer review intervention to minimize discrepancies between preregistrations and published manuscripts: a feasibility study. (2021) doi:10.17605/OSF.IO/5DH47.

205. Nosek, B. A. *et al.* Promoting an open research culture. *Science* **348**, 1422–1425 (2015).

206. Lilburn, S. D., Little, D. R., Osth, A. F. & Smith, P. L. Cultural problems cannot be solved with technical solutions alone. *Comput. Brain Behav.* **2**, 170–175 (2019).

207. Brainerd, C. J. & Reyna, V. F. Replication, registration, and scientific creativity. *Perspect. Psychol. Sci.* (2018) doi:10.1177/1745691617739421.

208. Appelbaum, M. *et al.* Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *Am. Psychol.* **73**, 3 (2018).

209. Simera, I., Moher, D., Hoey, J., Schulz, K. F. & Altman, D. G. A catalogue of reporting guidelines for health research. *Eur. J. Clin. Invest.* **40**, 35–53 (2010).

210. Haven, T. L. *et al.* Preregistering qualitative research: a Delphi study. *Int. J. Qual. Methods* **19**, 1–13 (2020).

211. Chambers, C. D., Feredoes, E., Muthukumaraswamy, S. D. & Etchells, P. J. Instead of "playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neurosci.* **1**, 4–17 (2014).

Supplementary Information

for

Reducing bias, increasing transparency, and calibrating confidence with preregistration

Tom E. Hardwicke & Eric-Jan Wagenmakers

Department of Psychology

University of Amsterdam

Contents

Supplementary Information A. Tutorials, templates, and case studies

Supplementary Information B. Registries

Supplementary Information A. Tutorials, templates, and case studies

- General quantitative research (Open Science Framework; Bowman et al., 2020)
- General quantitative psychology (PRP-QUANT; Bosnjak et al., 2021)
- Qualitative research (Haven et al., 2020)
- Social psychology (van 't Veer & Giner-Sorolla, 2016)
- Infant research (Havron et al., 2020)
- Psychopathology research (Krypotos et al., 2019)
- Secondary data analysis (Akker et al., 2019; Mertens & Krypotos, 2019)
- Event Related Potentials (Paul et al., 2021)
- Experimental linguistics (Roettger, 2019)
- Economics (Olken, 2015)
- Experience sampling research (Kirtley et al., 2021)
- Expertise research (Moreau, 2019)
- Cognitive modelling (Crüwell & Evans, 2019)
- Developmental economics (Casey et al., 2012)
- Political science (Humphreys et al., 2013)

Supplementary Information B. Registries

- Open Science Framework Registry (https://osf.io/registries/)
- Open Science Framework Metascience Registry (https://osf.io/registries/metascience/discover)
- ClinicalTrials.gov (https://clinicaltrials.gov/)
- European Union Clinical Trials Register (https://www.clinicaltrialsregister.eu/)
- American Economic Association RCT Registry (https://www.socialscienceregistry.org)
- Animal Study Registry https://www.animalstudyregistry.org
- Preclinicaltrials.edu https://preclinicaltrials.eu
- PROSPERO for systematic reviews (https://www.crd.york.ac.uk/prospero/)
- Registry for International Development Impact Evaluations (RIDIE; https://ridie.3ieimpact.org/)
- Evidence in Governance and Politics Registry (EGAP; https://egap.org/)
- Registry of Efficacy and Effectiveness Studies in education (https://sreereg.icpsr.umich.edu/sreereg/)

- World Health Organisation (WHO) International Clinical Trials Registry Network (https://www.who.int/clinical-trials-registry-platform) [NOTE: this is a collection of clinical trial registries that meet specific criteria outlined by WHO]
- AsPredicted (https://aspredicted.org/) [NOTE: because preregistrations logged with this website can be kept private, it cannot be considered a public registry. We mention it here for completeness, as it is a popular website for creating preregistrations].

Supplementary references

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

Bosnjak, M., Fiebach, C., Mellor, D. T., Mueller, S., O'Connor, D., Oswald, F., & Sokol-Chang, R. (2021). A template for preregistration of quantitative research in psychology: Report of the joint psychological societies preregistration task force. PsyArXiv. https://doi.org/10.31234/osf.io/d7m5r

Bowman, S. D., DeHaven, A. C., Errington, T. M., Hardwicke, T. E., Mellor, D. T., Nosek, B. A., & Soderberg, C. K. (2020). OSF Prereg Template [Preprint]. MetaArXiv. https://doi.org/10.31222/osf.io/epgjd

Casey, K., Glennerster, R., & Miguel, E. (2012). Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan*. The Quarterly Journal of Economics, 127(4), 1755–1812. https://doi.org/10.1093/qje/qje027

Crüwell, S., & Evans, N. J. (2019). Preregistration in complex contexts: A preregistration template for the application of cognitive models. https://doi.org/10.31234/osf.io/2hykx

Haven, T. L., Errington, T. M., Gleditsch, K., van Grootel, L., Jacobs, A. M., Kern, F., Piñeiro, R., Rosenblatt, F., & Mokkink, L. (2020). Preregistering qualitative research: A Delphi study. International Journal of Qualitative Methods, 19, 1–13. https://doi.org/10.31235/osf.io/pz9jr

Haven, T. L., & Grootel, D. L. V. (2019). Preregistering qualitative research. Accountability in Research, 26(3), 229–244. https://doi.org/10.1080/08989621.2019.1580147

Havron, N., Bergmann, C., & Tsuji, S. (2020). Preregistration in infant research—A primer. Infancy, 25(5), 734–754. https://doi.org/10.1111/infa.12353

Humphreys, M., Sierra, R. S. de la, & Windt, P. van der. (2013). Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration. Political Analysis, 21(1), 1–20. https://doi.org/10.1093/pan/mps021

Kirtley, O. J., Lafit, G., Achterhof, R., Hiekkaranta, A. P., & Myin-Germeys, I. (2021). Making the black box transparent: A template and tutorial for registration of studies using experience-sampling methods. Advances in Methods and Practices in Psychological Science, 4(1), 2515245920924686. https://doi.org/10.1177/2515245920924686

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(6), 517–527. https://doi.org/10.1037/abn0000424

Mertens, G., & Krypotos, A.-M. (2019). Preregistration of Analyses of Preexisting Data. Psychologica Belgica, 59(1), 338–352. https://doi.org/10.5334/pb.493

Moreau, D. (2019). Preregistration in the context of expertise research: Benefits, challenges, and recommendations. PsyArXiv. https://doi.org/10.31234/osf.io/v7xrb

Olken, B. A. (2015). Promises and perils of pre-analysis plans. Journal of Economic Perspectives, 29(3), 61–80. https://doi.org/10.1257/jep.29.3.61

Paul, M., Govaart, G. H., & Schettino, A. (2021). Making ERP research more transparent: Guidelines for preregistration. International Journal of Psychophysiology. https://doi.org/10.1016/j.ijpsycho.2021.02.016

Roettger, T. B. (2019). Preregistration in experimental linguistics: Applications, challenges, and limitations. PsyArXiv. https://doi.org/10.31234/osf.io/vc9hu

Simera, I., Moher, D., Hoey, J., Schulz, K. F., & Altman, D. G. (2010). A catalogue of reporting guidelines for health research. European Journal of Clinical Investigation, 40(1), 35–53. https://doi.org/10.1111/j.1365-2362.2009.02234.x

van ’t Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. Journal of Experimental Social Psychology, 67, 2–12. https://doi.org/10.1016/j.jesp.2016.03.004