

## Methodology and Research Practice

# When and How to Deviate From a Preregistration

Daniël Lakens<sup>1a</sup>

<sup>1</sup> IEIS, Eindhoven University of Technology, Eindhoven, The Netherlands

Keywords: preregistration, open science, severity, falsificationism, statistical analysis plan

<https://doi.org/10.1525/collabra.117094>

---

## Collabra: Psychology

Vol. 10, Issue 1, 2024

---

As the practice of preregistration becomes more common, researchers need guidance in how to report deviations from their preregistered statistical analysis plan. A principled approach to the use of preregistration should not treat all deviations as problematic. Deviations from a preregistered analysis plan can both reduce and increase the severity of a test, as well as increase the validity of inferences. I provide examples of how researchers can present deviations from preregistrations and evaluate the consequences of the deviation when encountering 1) unforeseen events, 2) errors in the preregistration, 3) missing information, 4) violations of untested assumptions, and 5) falsification of auxiliary hypotheses. The current manuscript aims to provide a principled approach to deciding when to deviate from a preregistration and how to report deviations from an error-statistical philosophy grounded in methodological falsificationism. The goal is to help researchers reflect on the consequence of deviations from preregistrations by evaluating the test's severity and the validity of the inference.

Researchers increasingly create preregistrations, which are time-stamped documents that transparently communicate that analyses have not been selected based on information in the data that will determine the outcome of the planned analyses. Preregistrations should specify all decisions about how data will be analyzed, including information about the amount of data that will be collected, the measures that will be used, how data will be pre-processed, and when test results would corroborate or falsify a prediction. Across scientific disciplines, the practice of study preregistration emerged in response to concerns about the opportunistic use of flexibility when analyzing data in quantitative research (e.g., Chan et al., 2013; Wiseman et al., 2019). When researchers misuse analytic flexibility, they increase the probability that statistical tests will lead to a desired, but wrong, scientific claim. Such practices impact the *severity* of the test, which is high when 1) it is likely that a test will support a hypothesis when the hypothesis is correct, and 2) it is likely that a test will fail to support a hypothesis when the hypothesis is wrong (Mayo, 2018).

There is widespread agreement that scientific claims should be severely tested. Although not all researchers might be aware of it, when they make dichotomous claims based on whether  $p < \alpha$ , or when they perform an a-priori power analysis, they are putting a Neyman-Pearson error-statistical approach into practice (Neyman & Pearson, 1933). This Neyman-Pearson approach to hypothesis testing is based on a philosophy of science known as method-

ological falsificationism (Lakatos, 1978). Data is used to make scientific claims about whether a prediction was corroborated or not. By controlling the Type 1 and Type 2 error rates in the long run few claims will be incorrect. The lower both error rates are, the more severe the test (Popper, 1959). Preregistration allows readers of a scientific article to transparently evaluate the severity of reported tests (Lakens, 2019). It is a useful tool to identify problematic practices such as 'hypothesizing after results are known' (Kerr, 1998). When researchers develop a hypothesis after looking at the data and then present the data as a test of that hypothesis, the hypothesis might very well be true, but it has not been severely tested. After all, if the hypothesis was wrong there was no possibility that the data would fail to provide support for the hypothesis. Similarly, preregistration can help to identify problematic practices such as opportunistically reporting only those analyses that support the hypothesis, when much more defensible analyses do not support the hypothesis. These practices that reduce the severity of a test lead to *biasing selection effects* (Mayo, 2018). The presence of a preregistration will allow readers to independently verify that biasing selection effects are absent, and that researchers severely tested their hypothesis.

Although some researchers manage to report a study that was performed exactly in line with their preregistration, many researchers deviate from their preregistration when they perform their study and analyze their data

---

<sup>a</sup> Correspondence can be addressed to Daniël Lakens, Human-Technology Interaction Group, ATLAS 9.042, PO Box 513, 5600MB Eindhoven, The Netherlands. E-mail: D.Lakens@tue.nl

(Akker et al., 2023; Claesen et al., 2021; Heirene et al., 2021; Ofosu & Posner, 2023; TARG Meta-Research Group and Collaborators, 2023; Toth et al., 2021). Common reasons for deviations are collected sample sizes that do not match the preregistered sample size, excluding data from the analysis for reasons not prespecified, performing a different statistical test than preregistered, or implementing changes to the analysis plan due to errors during the data collection. Here I focus on scientific claims that are based on a statistical test that is not identical to the test in the preregistered statistical analysis plan. Researchers do not need to adhere to a preregistration when statistical tests are used to *generate* hypotheses. Deviations from preregistration are also distinct from sensitivity analyses, which are (often unsystematic) variations of statistical tests that are performed to reach a qualitative assessment of the extent to which a claim depends on analytic decisions.

To illustrate, a deviation from a preregistration would occur when researchers preregistered to analyze all data, but after inspection of the data decide to exclude a subset of the observations, and subsequently use the results of the analysis based on a subset of the data as the basis of their claim while the analysis they originally planned is ignored. When statistical tests are used to generate hypotheses researchers would perform the analysis on the subset of the data, but use the results of the analysis based on a complete dataset as the basis of their claim while presenting the idea that certain observations should be excluded when investigating the research question as a hypothesis that should be examined in a future study. In a sensitivity analysis researchers present both the analysis of the full dataset and a subset of the dataset and use the results of the analysis based on a complete dataset as the basis of their claim, while noting that the results are (or are not) dependent on decisions made during the data analysis. Note that I do not use the terms ‘exploratory’ or ‘confirmatory’ analysis, as from a methodological falsificationist philosophy of science the more relevant distinction is the severity of the statistical test, or whether error rates are controlled. A ‘non-confirmatory’ test can be more severe than a ‘confirmatory’ test. For example, researchers can preregister a test that does not control error rates, or researchers can deviate from a preregistration because a test assumption (such as the homogeneity assumption) is statistically rejected, and the non-preregistered test has better error control.

Researchers often perform many tests in a scientific article, but not all tests of a hypothesis lead to a scientific claim. According to Popper (1959) scientific claims are dichotomous statements regarding whether a prediction was supported or not (see also Frick, 1996; Uygun Tunç et al., 2023). The scientific claims based on hypothesis tests in a scientific literature determine how good the track record is of the theory that the predictions were derived from (Meehl, 1990). Specifically in the context of a test of a theory researchers need to specify which statistical test result they will interpret as support for their prediction, and which result they will interpret as falsifying their prediction (Lakens & DeBruine, 2021). Although the focus in this article is on how the error-statistical philosophy proposed by

Mayo provides a coherent justification for preregistration in frequentist hypothesis tests (Lakens, 2019), severe tests are deemed desirable in Bayesian statistics as well. For example, severity underlies the practice of model checking in Bayesian analysis (Gelman & Shalizi, 2013), it is central to assessing the testability of cognitive models based on the prior predictive distribution from Bayesian statistics (Vanpaemel, 2019), and Bayesian approaches to severity are also based on controlling the probability of misleading evidence (van Dongen et al., 2023).

At the same time, severe tests (or the prevention of biasing selection effects) is not the only goal when testing hypotheses. Sometimes researchers should deviate from preregistered statistical analyses because doing so increases the validity of the statistical inference. This article aims to provide a principled approach to reflecting on the consequence of deviations from preregistrations by evaluating the test’s severity and validity of the inference.

### Deviating from Preregistrations

The goal of a statistical hypothesis test in an error statistical philosophy is to make *valid claims* that are *severely tested* (Mayo & Spanos, 2011). In line with methodological falsificationism (Lakatos, 1978) the principle is that researchers build on theories that have made successful predictions based on severe tests (Meehl, 1990). Although alternative philosophical justification of preregistration have been proposed (e.g., Peikert et al., 2023), preregistration follows most naturally from a philosophy of science that values successful predictions based on tests that strictly control erroneous claims (Lakens, 2019). The goal of prespecifying a statistical analysis plan is to prevent biasing selection effects, which can increase the probability of an incorrect claim, and thereby the severity of the test.

One justifiable reason to deviate from a preregistered statistical analysis plan is to increase the validity of the scientific claim – even if doing so comes at the expense of the severity of the test. Validity refers to “the approximate truth of an inference” (Shadish et al., 2001, p. 34). When researchers can make a convincing argument that the preregistered analysis plan leads to a statistical test with low validity, a less severe but more valid test of the hypothesis might lead to a claim that has more verisimilitude, or truth-likeness (Niiniluoto, 1998; Popper, 1959). Both validity, which is a property of the inference, and severity, which is a property of the test (Claesen et al., 2022) are continuous dimensions. A statistical test can be more or less severe, and the inference can be more or less valid. It is important to note that in practice a claim based on a hypothesis test that contains a deviation from a preregistration will often be more severely tested than a claim based on a non-preregistered test. Such deviations should not just be reported, but the consequences of the deviation should also be evaluated (Hardwicke & Wagenmakers, 2023; Willroth & Atherton, 2024). [Table 1](#) provides four examples of tests with lower or higher severity and lower or higher validity.

When evaluating the severity of a test we do so irrespective of whether the scientific claim based on the test is true

**Table 1. Examples of reporting practices that lead to tests with higher or lower severity and claims with higher or lower validity.**

	Lower validity	Higher validity
Lower severity	Selectively reporting one out of five variables that measure a construct of interest because only this test yields $p < .05$ .	Deviating from a preregistration to exclude observations not caused by processes related to the research question.
Higher severity	Following a preregistered analysis of all data even though 15% of respondents did not follow the instructions.	Following a preregistered statistical analysis plan with high construct and statistical validity.

or not. It is perfectly possible to make true scientific claims based on tests that lack any severity. For example, a researcher can commit scientific fraud, fabricate data in support of a claim, only for subsequent rigorous research to corroborate the claim. Similarly, a researcher can hypothesize after results are known and pretend that they performed a study that tested this hypothesis, only for subsequent studies to corroborate their claim. As scientists, we can never know if a scientific claim is true or not (Popper, 1959). But we can evaluate how severely a claim has been tested. Although it will rarely be possible to precisely quantify severity, we can roughly evaluate the impact analytic decisions could have on the inflation of the Type 1 error rate (e.g., Stefan & Schönbrodt, 2023).

When researchers evaluate the validity of an inference, they consider aspects such as construct validity, external validity, internal validity, and statistical conclusion validity (Shadish et al., 2001; Vazire et al., 2022). It is never possible to know if inferences are valid, but the risk of threats to validity can be evaluated. Deviations from a preregistration mainly impact statistical conclusion validity (e.g., excluding data that is not generated by the mechanisms of interest) and construct validity (e.g., changing which items are included in a scale based on the reliability of the scale). External validity (e.g., claims about how effects generalize) and internal validity (e.g., confounds) are less directly influenced by deviations from a preregistered analysis plan, but could emerge due to errors (e.g., when a covariate is accidentally not stored during data collection). Importantly, statistical conclusion validity can also be increased when researchers deviate from a statistical analysis plan, for example when the preregistration contained errors or when more data is collected than planned.

In the remainder of this paper, I will differentiate between five categories of reasons to deviate from a preregistration based on a review of deviations reported in meta-scientific research (Akker et al., 2023; Claesen et al., 2021; Toth et al., 2021; Willroth & Atherton, 2024), as well as based on anecdotal experiences when peer reviewing preregistered studies or preregistering my research. Although perhaps not exhaustive, these categories encompass most of the reasons why researchers would report deviations from a preregistration. I will provide examples of how researchers can present deviations from preregistrations and evaluate the consequences of the deviation for the severity of the test and the validity of the inference. I will focus on 1) unforeseen events, 2) errors in the preregistration, 3) missing information, 4) violations of untested

assumptions, and 5) falsification of auxiliary hypotheses. These categories might overlap in practice, but they broadly differ in the extent to which they can reduce or increase the severity and validity. As summarized in Figure 1 each of the five categories of reasons for deviations requires researchers to specify the reason for the deviation and evaluate the consequences of the deviation for the severity of the test. For some deviations researchers can optionally discuss how the deviation led to an inference with higher validity.

### Unforeseen event

Researchers sometimes need to deviate from a preregistration due to an unforeseeable event that interferes with the data collection. During the COVID-19 pandemic, many researchers faced difficulties collecting data. Deviations from preregistrations were unavoidable because the desired number of in-person observations could not be collected, or because data collection had to be moved online. One quite common reason to deviate from a preregistration is that the planned sample size does not match the final sample size after data cleaning (Akker et al., 2023; Toth et al., 2021). Although the planned and collected sample size can also differ because of a simple error, or because assumptions were violated, often researchers simply do not foresee how difficult or easy it will be to collect data.

Unforeseen events are largely outside the control of a researcher. The less control researchers have over the final sample size, the less flexibility there is in the data analysis, and therefore there is often little risk of biasing selection effects, as long as data-dependent stopping rules are avoided. Researchers typically perform the preregistered analysis on the data, even if the sample size turns out to be larger or smaller than planned. If researchers can convincingly argue that they did not engage in optional stopping or selective reporting (Simmons et al., 2011) a deviation in the preregistered sample size does not inflate the Type 1 error rate. A smaller number of observations than planned can reduce the severity of the test, but primarily because the lower statistical power for the effect of interest makes it less likely that the test will support the hypothesis if the hypothesis is true. A larger sample size than preregistered, all else equal, increases the test's severity, as the test is more likely to support the prediction when it is true. When sample sizes are substantially smaller, inconclusive results might have become more likely, and researchers might need to re-evaluate which research question they can adequately

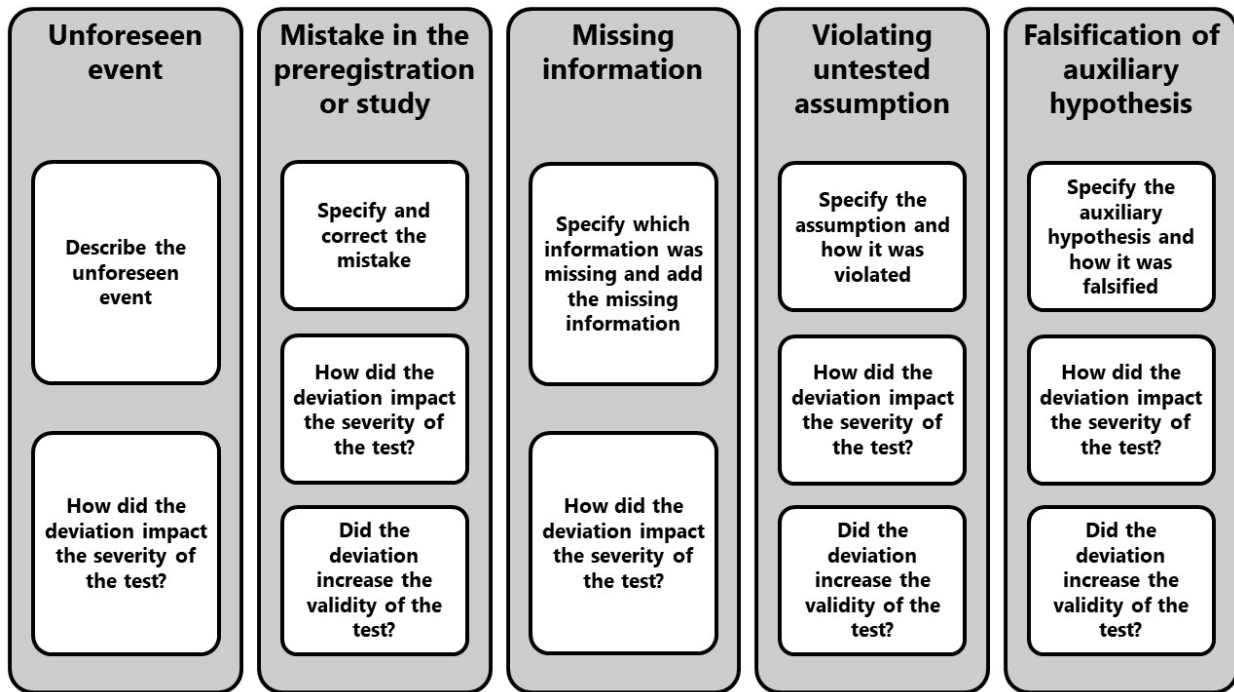


Figure 1. Overview of the five categories of reasons to deviate from a preregistration.

answer (Lakens, 2022). It might even make little sense to perform a hypothesis test if the statistical power is very low.

Given that researchers often collect a different number of observations as specified in the preregistration (Akker et al., 2023), one might argue researchers should always plan for some flexibility in the final number of observations they will analyze. It can be beneficial to design studies that allow researchers to repeatedly analyze the data in a group sequential analysis (Lakens, 2014; Proschan et al., 2006; Wassmer & Brannath, 2016). Importantly, the use of an alpha spending approach where a predetermined Type 1 error rate is spent at each interim look at the data allows researchers to preregister an analysis plan for an intended sample size, but also prespecify the approach that will be used to compute the alpha level when more or less data than planned is collected, while controlling the Type 1 error rate.

If unforeseen events occur before the data is analyzed researchers can update the preregistration to transparently communicate they did not engage in optional stopping (Simmons et al., 2011). The Open Science Framework offers users the option to update registrations. If a paper is submitted as a Registered Report, the changes required due to unforeseen events typically warrant contacting the handling editor. The editor (who might consult with the peer reviewers of the Stage 1 report) can decide whether the changes are substantial enough to withdraw the in principle acceptance. If handled transparently, a different sample size than preregistered would typically not impact the Type 1 error rate, or the validity of the test, beyond the reduction in statistical power.

When unforeseen events occur, researchers should 1) describe the unforeseen event, and how it impacted the preregistration, and 2) discuss the change to the preregistra-

tion, and evaluate the possibility that the deviation from the preregistered analysis plan reduces the severity of the test. A researcher might describe the following deviation from their preregistration if unforeseen events occurred that led to less data than planned:

1. Due to a fire in the building the lab was unavailable for the planned data collection after data from 100 participants was collected. We specified that we would collect 240 observations. As data collection required EEG measurements, and we did not have access to a lab with EEG equipment, we could not collect data from the remaining 140 participants in the time we had to complete this project. The final sample size is therefore substantially smaller than planned.
2. The data file shows EEG recordings were made up to the day before the fire in the lab. All data we collected is included in the final analysis. We performed the planned analysis, only on fewer participants. The Type 1 error rate is therefore controlled. To examine the impact on the Type 2 error rate we performed a sensitivity power analysis. With 100 participants instead of 240 our planned Type 2 error rate was increased from 10% (i.e., 90% power) to 45% (i.e., 55% power). This is a substantial reduction in the informational value of our study. Inconclusive results are more likely.

Alternatively, a researcher might describe the following deviation from their preregistration if they collected more data than planned:

1. In our preregistration we expected to collect data from 500 participants based on past experiences recruiting participants through social media. Unexpect-



edly, our request to collect data went viral, and in the end 5124 participants completed our survey. We only analyzed the data once and included all participants who completed the survey before we started the data analysis.

2. We performed the analysis as originally planned but had much higher power to detect the effect of interest. This deviation did not reduce the severity of the test but increased it.

### Mistake in the preregistration or study

Mistakes happen, and open science practices will make mistakes more visible (Bishop, 2018). There are distinct types of mistakes researchers can make in preregistrations and the subsequent data collection. Some mistakes are minor slips. A researcher might have incorrectly stated that they predicted that the experimental condition would score *higher* than the control condition on the dependent variable, while the theory clearly states that the experimental condition should score *lower*. Researchers might preregister their analysis plan in computer code based on a simulated dataset but specify the wrong variable name in the analysis. Alternatively, researchers might plan to collect data from 80 participants, and mistakenly recruit one participant too many (Claesen et al., 2021). Some mistakes might be due to the misapplication of a rule. For example, researchers might preregister that they will consider a prediction falsified if they observe a statistically significant difference. Although hypothesis tests are commonly used to claim there is a difference between conditions, the researchers realize during the data analysis that with 56,000 observations tiny and practically insignificant differences will be statistically significant. In hindsight, they believe the better test of their prediction would have been to test if there is a difference that is large enough to matter.

Some mistakes can be corrected in only one way. If all peers agree a theory predicts a lower score instead of a higher score, the correction is obvious, and the deviation from the preregistered analysis plan to correct the mistake does not noticeably impact the evaluation of the severity of the test. If one additional participant is erroneously collected the statistical claim might remain unchanged regardless of the data this additional participant provided.

Other mistakes might allow more flexibility in how they can be corrected. If researchers realize they relied on a null hypothesis significance test when their research question actually called for a test against a smallest effect size of interest it is less clear that they made a mistake. They might have chosen to rephrase their research question opportunistically, and they would have considerable flexibility to choose which effects are considered large enough to matter. In such cases there might be a substantial reduction in the severity of the test, especially when researchers have already seen the data. Nevertheless, peers might still agree that the scientific claim should be based on an analysis that deviates from the preregistration, and not on the preregistered analysis plan.

It is also possible that researchers are satisfied with their preregistered analysis, but that peer reviewers consider the

performed analysis to be a mistake. Common deviations in preregistrations are a switch from a simpler to a more complex statistical model, or vice versa (Akker et al., 2023). One reason these deviations are common might be that researchers often preregister an analysis plan that is not the best test of their research question. Furthermore, advances in statistics might mean there are better statistical approaches available than a researcher has preregistered. Although it is always possible to present multiple statistical approaches as a sensitivity analysis, the results might differ across the tests. This can lead to a discussion between researchers and reviewers about which test is most valid, and should be the basis of the scientific claim. The trade-off between the severity of the test and the validity of the inference might mean that deviating from the preregistration is the best decision, unless the deviation is driven by biasing selection effects. Submitting the preregistration for peer review, as is done in Registered Reports, can prevent discussions with peer reviewers after the data has been collected (Nosek & Lakens, 2014). After a single study there can remain considerable uncertainty about which analysis is most valid, and future replication and extension studies should be performed to examine this question.

When researchers realize they need to deviate from their preregistration due to a mistake, they should 1) specify which mistake was made, and correct the mistake, and 2) evaluate the possibility that the deviation from the preregistered analysis plan reduces the severity of the test. This requires evaluating how much flexibility researchers would have when correcting the mistake. Optionally, researchers might want to explain how the deviation increases the validity of the inference. For example, a researcher might describe the following deviation from their preregistration if they made a mistake in their preregistration:

1. We specified that the dependent variable would be computed by summing five items, but failed to specify the scores for items 2 and 5 should first be inverted.
2. As noted in the introduction, the dependent variable was used in 3 earlier studies. In all studies items 2 and 5 were reverse coded before summing the items. We believe that despite our mistake it is clear there is only one valid way to sum the five items, and therefore our deviation does not impact the severity of the test, but increases the validity of the inference.

Another example, where a variable is missing due to a programming error, could be communicated as follows:

1. Due to a programming error our dataset did not contain the responses from participants whether they currently owned a dog, or not. As our preregistered analysis plan involved including dog-ownership as a factor we could not analyze the data as preregistered. We therefore had to adjust our statistical analysis.
2. As the factor 'dog-ownership' could not be included in the analysis, it was removed. We cannot think of any other possibility to change the statistical analysis plan. Given that the absence of the factor dog-ownership did not provide us with any flexibility in the data analysis, the Type 1 error rate is not impacted, and

the deviation from our preregistered analysis plan does not reduce the severity of the test. However, an analysis that includes dog ownership as a factor would have been a more valid test of our prediction.

### Missing information

Sometimes researchers forget to specify information in a preregistration, or the preregistration is so vague about specific aspects of the design or analysis that the information might as well be treated as missing. Missing information is an error of omission. When a preregistration is missing information researchers need to make a decision about the statistical analyses they will perform at a time when they have access to the data. These decisions might formally not be a deviation from a preregistration, but researchers still need to specify how the final analysis deviated from the preregistered analysis plan. When information is missing there is a possibility that the statistical analysis plan is chosen opportunistically and reduces both the severity and validity of the test. This in no way means that every deviation from a preregistration is opportunistic, or even that every deviation from a preregistration reduces the severity of the test (Lakens, 2019).

The amount of flexibility that is available to opportunistically choose an analysis depends on which information is missing. For example, if researchers forgot to specify the alpha level of a single hypothesis test, and they add the missing information by declaring the hypothesis test will be performed at an alpha level of 0.05, we can safely assume that this is the value researchers would have prespecified had they not forgotten to include it in the first place. However, if researchers collected 5 dependent variables, and did not specify whether their prediction is corroborated when any of the 5 variables yields a significant result, or only if all variables yield a significant result, there is considerable variability in whether multiplicity corrections should be applied, and which correction for multiplicity is chosen. Note that we are not concerned with an evaluation of whether a researcher has opportunistically chosen an analysis, but with the extent to which they can transparently communicate the severity of the test.

Concerns regarding the possible impact of flexibility in the data analysis on the severity of the test can be reduced if researchers can show that all plausible analyses (and especially the most conservative test) lead to the same scientific claim. If the same statistical claim can be made for even the most conservative analysis, then the missing information in the preregistration was inconsequential for the severity of the test. If there is substantial flexibility due to the missing information not all plausible analyses might lead to the same statistical claim. If the statistical claims depend on analytic choices, researchers will need to communicate that the claim is less severely tested than if it had been adequately preregistered.

Deviations from a preregistration due to a lack of information can therefore at best maintain the severity of a test (e.g., when some information is lacking, but there is only a single plausible option for the analysis). In practice, it is quite likely that preregistrations that lack im-

portant information will reduce the severity of the test compared to preregistrations where this information was specified. For some missing information, such as the total amount of data that will be collected, it might not even be possible to perform all plausible alternative analyses (as this would include collecting more data, and analyzing data with data points that are unavailable because data collection was stopped earlier). The probability that information is missing can be reduced by preregistering the analysis code.

When researchers realize that they need to deviate from their preregistration due to missing information they should 1) specify which information was missing, and add the missing information, and 2) evaluate the possibility that the additions to the preregistered analysis plan reduce the severity of the test. This requires an evaluation of how much flexibility researchers had when adding the missing information. For example, a researcher might describe the following deviation from their preregistration if they failed to specify the direction of a *t*-test:

1. We failed to specify the directionality of the test but planned to perform a one-sided test.
2. Although we planned a one-sided test, it is also common to perform two-sided tests. If we had performed a two-sided test, the *p* value of 0.002 would still have been smaller than the alpha level, as  $0.002 \times 2 < 0.05$ . The severity of the test is therefore not impacted due to the missing information.

As another example, where researchers failed to specify which simple effects would be performed after an ANOVA (analysis of variance):

1. We specified that we would perform a one-way ANOVA to compare the three conditions, but failed to specify which simple effects between the three conditions would support or falsify our prediction, and how we would control for multiple comparisons. We expected condition A to be larger than B, and B to be larger than C. Testing this prediction directly in a single contrast effect yielded a *p* value of 0.028.
2. Multiple tests could be performed to compare the three groups (e.g.,  $A = B$ ,  $A \& B < C$ ). Our prediction is also not the only pattern of the difference between means that peers might find plausible. After correcting for multiple comparisons, the test result would no longer confirm our prediction. Therefore, our statistical claim was less severely tested than if it had been preregistered.

When reporting the deviation, researchers should include the results of all relevant statistical tests, plots of the data, or other quantitative information that will help readers evaluate the severity of the test after deviating from the preregistered analysis plan. This information is not presented in the hypothetical examples in this manuscript. Deviations should be clearly indicated in a manuscript, but in case of strict word limits, details can be made available in a supplemental file.

## Violating untested assumption

Researchers make assumptions about how participants will behave when they participate in a study. Although ideally all participants follow instructions and provide valid responses, in practice some responses should be excluded from the data analysis. If researchers had theoretical models of how participants would respond in their study, they could accurately decide which responses were outliers, and exclude responses that the model indicates are invalid from the statistical analysis. Such models would then function as auxiliary hypotheses concerning which responses are valid, and which not. Regrettably, such models typically do not exist, and therefore researchers need to rely on more subjective evaluations when deciding which responses are outliers (Leys et al., 2019).

Some ways in which participants could generate invalid responses can be predicted, and therefore included in a preregistration. But because it is difficult to preregister all possible ways in which participants might behave when they take part in a study, it remains possible that researchers will need to deviate from their analysis plan after looking at the data, and apply additional (or sometimes fewer) criteria to exclude data than those detailed in the preregistration (Weaver & Rehbein, 2022). Specifying these additional criteria provides the opportunity to choose analyses that confirm what researchers want to find, and therefore it is possible that these deviations reduce the severity of the test. Due to the lack of theoretical models, deciding when untested assumptions are violated necessarily remains quite subjective. Researchers will need to explore the data and make decisions based on their observations. This distinguishes untested assumptions from auxiliary hypotheses, which can often be formally tested (as explained below).

As examples of untested assumptions, researchers might assume that participants will rarely leave questions unanswered, answer questions as instructed, and complete a survey only once. When analyzing the data, they might find that all these assumptions are violated. With experience, some commonly encountered cases might be included in a preregistration, but additional reasons to deviate might emerge. The decision whether a participant or a set of observations needs to be included can often not be completely determined based on an auxiliary theory but requires partly subjective judgment. It is not that researchers could not create a theory to guide decisions about which responses are valid and which not, but beyond some general approaches to determine careless responding specific theories that apply at the level of a specific study are often not available, and it might not even be worth the time and effort to develop them.

Sticking with a preregistered analysis plan when assumptions about the data that participants will provide are violated can substantially reduce the validity of an inference. Deviating from a preregistered analysis plan when assumptions are violated can therefore increase the validity of the inference. At the same time, deviating from a preregistered analysis plan opens up the possibility of biasing selection effects if the chosen analysis inflates the Type 1

error rate. It is possible that the risk of inflating the Type 1 error rate is rather small, and the increase in validity is substantial.

When researchers realize they need to deviate from their preregistration due to a violated assumption, they should 1) specify the assumption, and specify how the assumption was violated, and 2) specify how the analysis was changed to deal with the violation of the assumption, and evaluate the possibility that the change in preregistered analysis plan reduces the severity of the test, and if relevant, increases the validity of the inference. This requires evaluating how much flexibility researchers would have to deal with a violation of an assumption. When multiple options to deal with a violation of an assumption are available, researchers can report analyses for all defensible choices. If there are only a few defensible options to deal with a violated assumption, and all analyses lead to the same claim, a deviation from a preregistered analysis plan can be argued to have no effect on the severity of the test.

For example, a researcher might describe the following deviation from their preregistration if an assumption was violated:

1. We assumed all entered responses would be valid. However, two participants answered '1' to all survey questions, and one responded '1' to 80% of the questions. An analysis of the time they spent on the survey indicated their survey completion times were lower than that of all other participants. As we used reverse coded items, straight lining is less likely to be due to benign reasons (Reuning & Plutzer, 2020).
2. We decided to update our analysis plan by screening all data for careless responders. We defined careless responding by 'straight lining' (i.e., giving the same response) in more than 75% of the survey questions. As a consequence of this deviation of our preregistered analysis plan, three participants who showed 'straight lining' in either 100% or 80% of their responses were not included in the data analysis. The most common option to deal with straight lining is to remove participants, although data could also be treated as missing at random. An analysis of all data yielded no statistically significant result. Both approaches (removing participants, and imputing missing data), yielded significant results. There is a risk of biasing selection effects, as we might not have thought of removing data due to 'straight lining' had the preregistered analysis yielded a significant result. We consider the analyses where straight liners are removed or treated as missing data more valid, but we also acknowledge these analyses are less severely tested than if we had preregistered a data screening procedure for careless responding.

## Falsification of auxiliary hypothesis

We never test a theory in isolation, but always include auxiliary hypotheses about the measures and instruments that are used in a study, and the conditions realized in the experiment (Hempel, 1966). For example, a necessary as-

sumption in close replication studies is the ‘ceteris paribus’ clause. The ceteris paribus clause is an auxiliary hypothesis stating that all relevant factors required to observe a theoretically predicted effect are the same across different studies. For example, one study might provide written instructions in a serif font, while the replication study is the same in every way, except that it used a sans serif font. If theoretically the type of font of the instructions should not impact the effect size, font size can be relegated to the ceteris paribus clause, and the replication study does not differ in any meaningful way from the original. Auxiliary hypotheses can be systematically examined through replication and extension studies, and the ceteris paribus clause specifically can be examined through close replication studies, or ‘stability probes’ (Uygun Tunç & Tunç, 2022).

Researchers rely on auxiliary hypotheses concerning the reliability of measures, the effects of manipulations and instructions, and the validity of the method, design, and analysis. For example, a researcher might rely on the auxiliary hypothesis that the dependent variable will yield data that is normally distributed, because they plan to perform a statistical test that relies on the normality assumption. Such an auxiliary hypothesis is quite similar to a violation of an untested assumption, but the crucial difference is that the hypothesis can be tested based on the data that is collected. For example, a researcher might use the Jarque-Berra test to determine if the auxiliary hypothesis of normally distributed data can be rejected. If so, the researcher will perform a test that is robust against violations of the normality assumption. In a high-quality preregistration researchers can preregister conditional analyses (if X, then Y, else Z), which allow researchers to perform data-dependent analyses without impacting the severity of the test. If an auxiliary assumption is falsified, it no longer makes sense to report the preregistered analysis. This is different from violations of untested assumptions, where the assumptions underlying the preregistered analysis are never rejected outright.

Researchers also rely on the auxiliary hypothesis that the measure they use is sufficiently reliable. The reliability of a measure can be computed, for example based on the coefficient omega (Dunn et al., 2014) and compared to a minimum threshold that is considered satisfactory. If the reliability of a measure is low, alternative analysis strategies can be developed (e.g., by removing a bad item from a scale). Sometimes the falsification of an auxiliary hypothesis will mean that the intended hypothesis test should not be performed at all. For example, if researchers assume that participants will pay attention to task instructions, but almost all participants fail an attention check, or if a manipulation check indicates that a manipulation did not have the intended effect, the preregistered hypothesis test is not valid (Fiedler et al., 2021). Studies with failed attention checks or manipulation checks provide useful insights about the methodology that should be used in future studies, and might generate knowledge about the effectiveness of manipulations across different research settings, but if a test lacks validity, the inference based on the test does not inform us about whether or not the hypothesis is correct. If

participants did not read task instructions, or if a manipulation did not have the intended effect, it is highly probable that a null result will be observed, even if the theoretical prediction is in fact true.

When researchers realize they need to deviate from their preregistration due to a falsified auxiliary hypothesis, they should 1) specify the auxiliary hypothesis and report the methodological procedure that falsified the auxiliary hypothesis, and 2) specify how the analysis was changed to deal with the falsified auxiliary hypothesis, and evaluate whether the change in preregistered analysis plan reduces or increases the severity of the test, and if relevant, the effect on the validity of the inference. This requires evaluating how much flexibility researchers would have to deal with the falsified auxiliary hypothesis, how severe the test would be after a deviation from the preregistration, and the increase in validity when the statistical analysis plan is updated. If there are only a few defensible options to deal with falsified auxiliary hypotheses, all alternative analyses lead to the same claim, and deviating from the preregistration increases the validity of the inference, a deviation from a preregistered analysis plan is justified.

For example, a researcher might describe the following deviation from their preregistration if an auxiliary hypothesis was falsified:

1. Our preregistered analysis contained statistical tests that relied on the homogeneity assumption, or equal variances in both conditions. After computing the standard deviations, we noticed a surprisingly large difference, and Levene’s test and Bartlett’s test rejected the homogeneity assumption. It is possible that the planned Student’s *t*-test would insufficiently control the Type 1 error rate.
2. Based on recommendations to perform Welch’s *t*-test to control the Type 1 error rate when the homogeneity assumption is violated (Delacre et al., 2017), we deviated from the preregistered analysis plan and performed Welch’s *t*-test instead of Student’s *t*-test. We believe this deviation increases the statistical conclusion validity and the severity of our test as Welch’s *t*-test has better error control when the homogeneity assumption is violated, and not deviating could inflate the Type 1 error rate. There are several alternative options that could be performed when the assumption of homogeneity is violated, such as performing non-parametric tests. In our case, all of the commonly recommended non-parametric tests lead to the same statistical claim. Therefore, we believe the change in the analysis is justified, and the severity of the test is increased.

## Discussion

As the practice of preregistration becomes more common, researchers need guidance in how to report deviations from their preregistered statistical analysis plan due to unforeseen events, mistakes in the preregistration, missing information, violations of assumptions, or the falsification of auxiliary hypotheses. Based on the idea that the goal



of preregistration is to allow others to transparently evaluate the severity of a test, deviations from a preregistration should contain a reflection on the consequences of the deviation for the severity of the test. Some deviations have no impact on the test severity, while others decrease the severity substantially, even if they are often still more severe than non-preregistered tests. Under some circumstances deviating from a preregistration can increase the severity of a test. It can also be justified to deviate from a preregistration if doing so increases the validity of the inference.

Even though deviations from preregistrations should be expected, ideally, they are not needed. Some researchers do not need to deviate from their preregistration when analyzing the data (Akker et al., 2023; Claesen et al., 2021; Heirene et al., 2021). Sometimes deviations to the preregistration can be made before the data is available. Updating a preregistration can transparently communicate that these decisions were unlikely to suffer from biasing selection effects, as the data was not available when these decisions were made. Updating the preregistration has no consequences for the severity of a test, as long as the updates are independent of the data that is used to test the hypothesis. The more experience you have in preregistering hypotheses, and the more knowledge you have about the data you are collecting, the lower the probability that you will need to deviate from your preregistration. Collecting pilot data is one way to increase the knowledge you have about the data you will collect.

Researchers can preregister analysis code based on simulated data, or even create machine readable hypothesis tests that prevent any ambiguity or missing information (Lakens & DeBruine, 2021). Researchers can also use checklists to reduce the probability that information is missing from a preregistration (e.g., Wicherts et al., 2016). Perhaps even more important, researchers should not rush to a hypothesis test when they have skipped the necessary exploratory work that provides the input for a meaningful hypothesis test (Scheel et al., 2021). The less experience researchers have with the methods and measures they use in their study, the more likely it is that deviations will occur. In those cases, it can be useful to preregister conditional analyses or to use methods that allow for more analytic flexibility while controlling the Type 1 error rate, and therefore maintain the severity of the test.

Deviations from preregistrations should be carefully and transparently communicated. Forms to report deviations are available, typically as tables where each deviation is a row, and aspects of the deviation are described in columns<sup>1</sup>. It is useful to create preregistrations with line numbers to point back to specific sentences in the preregistration and refer to line numbers in the analysis code where deviations were introduced. Such a practice would also prevent researchers from forgetting to report preregistered analyses. For every deviation clearly specify when, where, and why

a deviation from a preregistration occurred, followed by an evaluation of the impact of the deviation on the severity of the test (and where relevant, the validity of the inference).

Whether deviations from a preregistration that claim to increase the validity of the inference actually do so can only be tested in follow-up studies. If deviations increase the validity of an inference at the expense of the severity of the test, then subsequent replication and extension studies should corroborate predictions when the updated analysis plan is used on new data. This provides an indication the change from the preregistration led to a progressive research line (Lakatos, 1978; Meehl, 1990). However, if deviations are due to biasing selection effects, subsequent studies will be less likely to corroborate predictions, and researchers enter a degenerative research line. Future meta-scientific research should examine how well researchers and peers can evaluate the validity of an inference (Schivone & Vazire, 2023). In the short-term, researchers who deviate from a preregistration can demonstrate the deviation is part of a progressive research line by performing a successful direct replication study.

Meta-scientific research has shown that deviations are common. At the same time, for many deviations the impact on the severity of the test will be negligible, and sometimes deviations can even increase the test's severity. Where the deviations reduce the severity of the test more strongly, researchers should be able to justify deviations by providing arguments for a substantial increase in the validity of the inference. A principled approach to the use of preregistration should not treat all deviations as problematic. Deviations from a preregistered analysis plan can both reduce and increase the severity of a test, and researchers need to carefully evaluate both the severity and the validity of scientific claims.

## Competing Interests

I have no competing interests to declare.

## Funding

This work was funded with the support of the Ammodo Science Award 2023 for Social Sciences.

## Acknowledgements

This manuscript is in part inspired by Quantitude episode [S3E07](#) "In Defense of Researcher Degrees of Freedom" and Within & Between episode [S4E11](#) "Revisiting Preregistration and Registered Reports". I would like to thank Smriti Mehta and Lisa Spitzer for useful feedback on an earlier version.

<sup>1</sup> [https://web.archive.org/web/20240310134055/https://docs.google.com/spreadsheets/d/1lbUj\\_8C0gob63wv\\_6QLsQunmzNh9Ig3L/edit#gid=381582667](https://web.archive.org/web/20240310134055/https://docs.google.com/spreadsheets/d/1lbUj_8C0gob63wv_6QLsQunmzNh9Ig3L/edit#gid=381582667), <https://web.archive.org/web/20240310134249/https://docs.google.com/document/d/1m7k53z38w18AJe56ucf-tunnHuFM7wDIMFjpoGenwN6k/edit>, <https://osf.io/6fk87>, or <https://osf.io/yrvcg>.

Submitted: January 10, 2024 PDT, Accepted: March 22, 2024

PDT

Downloaded from [http://online.ucpress.edu/collabra/article-pdf/10/1/117094/817387/collabra\\_2024\\_10\\_1\\_117094.pdf](http://online.ucpress.edu/collabra/article-pdf/10/1/117094/817387/collabra_2024_10_1_117094.pdf) by guest on 25 July 2024



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

## References

- Akker, O. van den, Bakker, M., Assen, M. A. L. M. van, Pennington, C. R., Verweij, L., Elsherif, M., Claesen, A., Gaillard, S. D. M., Yeung, S. K., Frankenberger, J.-L., Krautter, K., Cockcroft, J. P., Kreuer, K. S., Evans, T. R., Heppel, F., Schoch, S. F., Korbmacher, M., Yamada, Y., Albayrak-Aydemir, N., & Wicherts, J. (2023). The effectiveness of preregistration in psychology: Assessing preregistration strictness and preregistration-study consistency. In *MetaArXiv*. <https://doi.org/10.31222/osf.io/h8xjw>
- Bishop, D. V. M. (2018). Fallibility in Science: Responding to Errors in the Work of Oneself and Others. *Advances in Methods and Practices in Psychological Science*, 1(3), 432–438. <https://doi.org/10.1177/2515245918776632>
- Chan, A.-W., Tetzlaff, J. M., Gøtzsche, P. C., Altman, D. G., Mann, H., Berlin, J. A., Dickersin, K., Hróbjartsson, A., Schulz, K. F., Parulekar, W. R., Krleža-Jerić, K., Laupacis, A., & Moher, D. (2013). SPIRIT 2013 explanation and elaboration: Guidance for protocols of clinical trials. *BMJ*, 346, e7586. <https://doi.org/10.1136/bmj.e7586>
- Claesen, A., Gomes, S., Tuerlinckx, F., & Vanpaemel, W. (2021). Comparing dream to reality: An assessment of adherence of the first generation of preregistered studies. *Royal Society Open Science*, 8(10), 211037. <https://doi.org/10.1098/rsos.211037>
- Claesen, A., Lakens, D., Vanpaemel, W., & van Dongen, N. (2022). Severity and Crises in Science: Are We Getting It Right When We're Right and Wrong When We're Wrong? In *PsyArXiv*. <https://doi.org/10.31234/osf.io/ekhc8>
- Delacre, M., Lakens, D., & Leys, C. (2017). Why Psychologists Should by Default Use Welch's t-test Instead of Student's t-test. *International Review of Social Psychology*, 30(1). <https://doi.org/10.5334/irsp.82>
- Dunn, T. J., Baguley, T., & Brunsten, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology (London, England: 1953)*, 105(3), 399–412. <https://doi.org/10.1111/bjop.12046>
- Fiedler, K., McCaughey, L., & Prager, J. (2021). Quo Vadis, Methodology? The Key Role of Manipulation Checks for Validity Control and Quality of Science. *Perspectives on Psychological Science*, 1745691620970602. <https://doi.org/10.1177/1745691620970602>
- Frick, R. W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods*, 1(4), 379–390. <https://doi.org/10.1037/1082-989X.1.4.379>
- Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 66(1), 8–38. <https://doi.org/10.1111/j.2044-8317.2011.02037.x>
- Hardwicke, T. E., & Wagenmakers, E.-J. (2023). Reducing bias, increasing transparency and calibrating confidence with preregistration. *Nature Human Behaviour*, 7(1), Article1. <https://doi.org/10.1038/s41562-022-01497-2>
- Heirene, R., LaPlante, D., Louderback, E. R., Keen, B., Bakker, M., Serafimovska, A., & Gainsbury, S. M. (2021). Preregistration specificity & adherence: A review of preregistered gambling studies & cross-disciplinary comparison. *PsyArXiv*. <https://doi.org/10.31234/osf.io/nj4es>
- Hempel, (1966). *Philosophy of Natural Science*. Pearson.
- Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, 2(3), 196–217. [https://doi.org/10.1207/s15327957pspr0203\\_4](https://doi.org/10.1207/s15327957pspr0203_4)
- Lakatos, I. (1978). *The methodology of scientific research programmes: Volume 1: Philosophical papers*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511621123>
- Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology*, 44(7), 701–710. <https://doi.org/10.1002/ejsp.2023>
- Lakens, D. (2019). The value of preregistration for psychological science: A conceptual analysis. *Japanese Psychological Review*, 62(3), 221–230. [https://doi.org/10.24602/sjpr.62.3\\_221](https://doi.org/10.24602/sjpr.62.3_221)
- Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, 8(1), 33267. <https://doi.org/10.1525/collabra.33267>
- Lakens, D., & DeBruine, L. M. (2021). Improving Transparency, Falsifiability, and Rigor by Making Hypothesis Tests Machine-Readable. *Advances in Methods and Practices in Psychological Science*, 4(2), 2515245920970949. <https://doi.org/10.1177/2515245920970949>
- Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to Classify, Detect, and Manage Univariate and Multivariate Outliers, With Emphasis on Pre-Registration. *International Review of Social Psychology*, 32(1), 5. <https://doi.org/10.5334/irsp.289>
- Mayo, D. G. (2018). *Statistical inference as severe testing: How to get beyond the statistics wars*. Cambridge University Press. <https://doi.org/10.1017/9781107286184>
- Mayo, D. G., & Spanos, A. (2011). Error statistics. *Philosophy of Statistics*, 7, 152–198. <https://doi.org/10.1016/B978-0-444-51862-0.50005-8>
- Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1(2), 108–141. [https://doi.org/10.1207/s15327965pli0102\\_1](https://doi.org/10.1207/s15327965pli0102_1)

- Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 231(694–706), 289–337. <https://doi.org/10.1098/rsta.1933.0009>
- Niiniluoto, I. (1998). Verisimilitude: The Third Period. *The British Journal for the Philosophy of Science*, 49, 1–29. <https://doi.org/10.1093/bjps/49.1.1>
- Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, 45(3), 137–141. <https://doi.org/10.1027/1864-9335/a000192>
- Ofosu, G. K., & Posner, D. N. (2023). Pre-Analysis Plans: An Early Stocktaking. *Perspectives on Politics*, 21(1), 174–190. <https://doi.org/10.1017/S1537592721000931>
- Peikert, A., Ernst, M. S., & Brandmaier, A. M. (2023). Why does preregistration increase the persuasiveness of evidence? A Bayesian rationalization. <https://doi.org/10.31234/osf.io/cs8wb>
- Popper, K. R. (1959). *The logic of scientific discovery*. Routledge. <https://doi.org/10.1063/1.3060577>
- Proschan, M. A., Lan, K. K. G., & Wittes, J. T. (2006). *Statistical monitoring of clinical trials: A unified approach*. Springer.
- Reuning, K., & Plutzer, E. (2020). Valid vs. Invalid Straightlining: The Complex Relationship Between Straightlining and Data Quality. *Survey Research Methods*, 14(5), Article5. <https://doi.org/10.18148/srm/2020.v14i5.7641>
- Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspectives on Psychological Science*, 16(4), 744–755. <https://doi.org/10.1177/1745691620966795>
- Schiavone, S. R., & Vazire, S. (2023). Reckoning With Our Crisis: An Agenda for the Field of Social and Personality Psychology. *Perspectives on Psychological Science*, 18(3), 710–722. <https://doi.org/10.1177/17456916221101060>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Stefan, A. M., & Schönbrodt, F. D. (2023). Big little lies: A compendium and simulation of p-hacking strategies. *Royal Society Open Science*, 10(2), 220346. <https://doi.org/10.1098/rsos.220346>
- TARG Meta-Research Group and Collaborators. (2023). Estimating the prevalence of discrepancies between study registrations and publications: A systematic review and meta-analyses. *BMJ Open*, 13(10), e076264. <https://doi.org/10.1136/bmjopen-2023-076264>
- Toth, A. A., Banks, G. C., Mellor, D., O’Boyle, E. H., Dickson, A., Davis, D. J., DeHaven, A., Bochantin, J., & Borns, J. (2021). Study Preregistration: An Evaluation of a Method for Transparent Reporting. *Journal of Business and Psychology*, 36(4), 553–571. <https://doi.org/10.1007/s10869-020-09695-3>
- Uygun Tunç, D., & Tunç, M. N. (2022). A Falsificationist Treatment of Auxiliary Hypotheses in Social and Behavioral Sciences: Systematic Replications Framework. In *Meta-Psychology*. <https://doi.org/10.31234/osf.io/pdm7y>
- Uygun Tunç, D., Tunç, M. N., & Lakens, D. (2023). The epistemic and pragmatic function of dichotomous claims based on statistical hypothesis tests. *Theory & Psychology*, 09593543231160112. <https://doi.org/10.1177/09593543231160112>
- van Dongen, N., Sprenger, J., & Wagenmakers, E.-J. (2023). A Bayesian perspective on severity: Risky predictions and specific hypotheses. *Psychonomic Bulletin & Review*, 30(2), 516–533. <https://doi.org/10.3758/s13423-022-02069-1>
- Vanpaemel, W. (2019). The Really Risky Registered Modeling Report: Incentivizing Strong Tests and HONEST Modeling in Cognitive Science. *Computational Brain & Behavior*, 2(3–4), 218–222. <https://doi.org/10.1007/s42113-019-00056-9>
- Vazire, S., Schiavone, S. R., & Bottesini, J. G. (2022). Credibility Beyond Replicability: Improving the Four Validities in Psychological Science. *Current Directions in Psychological Science*, 31(2), 162–168. <https://doi.org/10.1177/09637214211067779>
- Wassmer, G., & Brannath, W. (2016). *Group Sequential and Confirmatory Adaptive Designs in Clinical Trials*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-32562-0>
- Weaver, E. J., & Rehbein, S. T. (2022). *Durchgeführt wie geplant? Ein detaillierter Vergleich zwischen Studien und ihren prä-registrierten Plänen*. <https://psycharchives.org/en/item/2462b05e-5d58-426b-8b43-a26556294a32>
- Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., Aert, V., M, R. C., Assen, V., & M, M. A. L. (2016). Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Frontiers in Psychology*, 7, 7. <https://doi.org/10.3389/fpsyg.2016.01832>
- Willroth, E. C., & Atherton, O. E. (2024). Best Laid Plans: A Guide to Reporting Preregistration Deviations. *Advances in Methods and Practices in Psychological Science*. <https://doi.org/10.31234/osf.io/dwx69>
- Wiseman, R., Watt, C., & Kornbrot, D. (2019). Registered reports: An early example and analysis. *PeerJ*, 7, e6232. <https://doi.org/10.7717/peerj.6232>

## Supplementary Materials

### Peer Review and Communication History

Download: [https://collabra.scholasticahq.com/article/117094-when-and-how-to-deviate-from-a-preregistration/attachment/225561.docx?auth\\_token=FhKN-0vJgR7S\\_kAj\\_ZAn](https://collabra.scholasticahq.com/article/117094-when-and-how-to-deviate-from-a-preregistration/attachment/225561.docx?auth_token=FhKN-0vJgR7S_kAj_ZAn)

---