



Thinking about evidence, and vice versa

[Menu](#)

---

[HOME](#)   [TABLE OF CONTENTS](#)   [FEEDBACK POLICY](#)   [SEMINAR](#)   [ABOUT](#)

---

[Menu](#)

---

## [64] How To Properly Preregister A Study

Posted on November 6, 2017 by Joe Leif Uri

P-hacking, the selective reporting of statistically significant analyses, continues to threaten the integrity of our discipline. *P*-hacking is inevitable whenever (1) a researcher hopes to find evidence for a particular result, (2) there is ambiguity about how *exactly* to analyze the data, and (3) the researcher does not perfectly plan out his/her analysis in advance. Although some mistakenly believe that accusations of *p*-hacking are tantamount to accusations of cheating, the truth is that accusations of *p*-hacking are nothing more than accusations of imperfect planning.

The best way to address the problem of imperfect planning is to plan more perfectly: to preregister your studies. Preregistrations are time-stamped documents in which researchers specify exactly how they plan to collect their data and to conduct their key confirmatory analyses. The goal of a preregistration is to make it easy to distinguish between planned, confirmatory analyses – those for which statistical significance is meaningful – and unplanned exploratory analyses – those for which statistical significance is not meaningful [1]. Because a good preregistration prevents researchers from *p*-hacking, it also protects them from suspicions of *p*-hacking [2].

In the past five years or so, preregistration has gone from being something that no psychologists did to something that many psychologists are doing. In our view, this wonderful development is the biggest reason to be optimistic about the future of our discipline.

But if preregistration is going to be the solution, then we need to ensure that it is done right. After casually reviewing several recent preregistration attempts in published papers, we noticed that there is room for improvement. We saw two kinds of problems.

### **Problem 1. Not enough information**

For example, we saw one “preregistration” that was simply a time-stamped abstract of the project; it contained almost no details about how data were going to be collected and analyzed. Others failed to specify one or more critical aspects of the analysis: sample size, rules for exclusions, or how the dependent variable would be scored (in a case for which there were many ways to score it). These preregistrations are time-stamped, but they lack the other critical ingredient: precise planning.

To decide which information to include in your preregistration, it may be helpful to imagine a skeptical reader of your paper. Let’s call him Leif. Imagine that Leif is worried that p-hacking might creep into the analyses of even the best-intentioned researchers. The job of your preregistration is to set Leif’s mind at ease [3]. This means identifying all of the ways you could have p-hacked – choosing a different sample size, or a different exclusion rule, or a different dependent variable, or a different set of controls/covariates, or a different set of conditions to compare, or a different data transformation – and including all of the information that lets Leif know that these decisions were set in stone in advance. In other words, your job is to prevent Leif from worrying that you tried to run your critical analysis in more than one way.

This means that your preregistration needs to be sufficiently exhaustive and sufficiently specific. If you say, “We will exclude participants who are distracted,” Leif could think, “Right, but distracted how? Did you define “distracted” in advance?” It is better to say, “We will exclude participants who incorrectly answered at least 2 out of our 3 comprehension checks.” If you say, “We will measure happiness,” Leif could think, “Right, but aren’t there a number of ways to measure it? I wonder if this was the only one they tried or if it was just the one they most wanted to report after the data came in?” So it’s better to say, “Our dependent variable is happiness, which we will measure by asking people ‘How happy do you feel right now?’ on a scale ranging from 1 (not at all happy) to 7 (extremely happy).”

If including something in a preregistration would make Leif less likely to wonder whether you p-hacked, then include it.

## Problem 2. Too much information

A preregistration cannot allow readers and reviewers to distinguish between confirmatory and exploratory analyses if it is not easy to read or understand. Thus, a preregistration needs to be easy to read and understand. This means that it should contain only the information that is essential for the task at hand. We have seen many preregistrations that are just too long, containing large sections on theoretical background and on exploratory analyses, or lots of procedural details that on the one hand will definitely be part of the paper, and on the other, are not p-hackable. Don't forget that you will publish the paper also, not just the preregistration; you don't need to say in the preregistration everything that you will say in the paper. A hard-to-read preregistration makes preregistration less effective [4].

To decide which information to exclude in your preregistration, you can again imagine that a skeptical Leif is reading your paper, but this time you can ask, "If I leave this out, will Leif be more concerned that my results are attributable to p-hacking?"

For example, if you leave out the literature review from your preregistration, will Leif now be more concerned? Of course not, as your literature review does not affect how much flexibility you have in your key analysis. If you leave out how long people spent in the lab, how many different RAs you are using, why you think your hypothesis is interesting, or the description of your *exploratory* analyses, will Leif be more concerned? No, because none of those things affect the fact that your analyses are confirmatory.

If excluding something from a preregistration would not make Leif more likely to wonder whether you p-hacked, then you should exclude it.

## The Takeaway

Thus, a good preregistration needs to have two features:

1. It needs to specify *exactly* how the key confirmatory analyses will be conducted.
2. It needs to be short and easy to read.

We designed [AsPredicted.org](https://aspredicted.org) with these goals in mind. The website poses a standardized set of questions asking you only to include what needs to be included, thus also making it obvious what does not need to be. The OSF offers lots of flexibility, but they also offer an AsPredicted template here: <https://osf.io/fnsb/> [5].

Still, even on AsPredicted, it is possible to get it wrong, by, for example, not being specific enough in your answers to the questions it poses. This table provides an example of how to wrongly and properly answer these questions.

Item in preregistration	<i>Bad answer</i>	<i>What's wrong with it?</i>	<i>Good answer</i>
<b>Research Question or Hypothesis</b>	Building on the work of Picasso (1901-1904), we hypothesized that....	You don't need reasons for asking the research question because they do not inform possible p-hacking. Just state the question or hypothesis of interest.	Question: Does sadness increase preference for the color blue?
<b>Dependent variable</b>	Preference for the color blue	This preference can be measured in many different ways so this statement underspecifies how it will be measured.	Participants will rate their liking for red, blue, orange, and purple on 7-point scales (1 = not at all; 7 = an extreme amount). Preference for blue will be defined as the difference between a participant's rating for blue and their average rating of the three non-blue colors.
<b>Manipulations/Conditions</b>	We will manipulate mood by having participants watch different videos.	This leaves room for cherry-picking from amongst a larger set of conditions. Specify the exact conditions and the exact manipulations.	Before rating their color preferences, participants will be randomly assigned to one of three conditions in which they watch a clip from either a sad video (My Dog Skip), a happy video (Pitch Perfect), or a neutral video (Gone With the Wind) .
<b>Analyses</b>	We will regress preference for the color blue on mood condition	There are many ways to run these analyses. For example, are you including covariates? How will "mood condition" be coded? If applicable, how will the standard errors be computed?	We will run an OLS regression predicting preference for the color blue with condition (coded 1 = sad video; 0 = happy or neutral video). We will control for gender (1 = male; 0 = female) in this analysis.
<b>Outliers &amp; Exclusions</b>	We will exclude participants who are inattentive, and those who show an extreme preference for the color orange.	What counts as "inattentive"? What counts as "extreme preference for the color orange"? You must define these things.	We will exclude participants who fail at least two out of the three attention checks that we will include at the beginning of our study (before the manipulation). We will also exclude participants whose rating of orange is higher than 5 on the 7-point scale.
<b>Sample size</b>	We conducted a power analysis that showed that ... And so we decided to collect between 100 and 200 observations.	Your power analysis is irrelevant to whether you p-hacked; leave it out. Also, any sample size between 100 and 200 is consistent with this preregistration.	We will stop data collection once 150 participants have submitted a response on MTurk. Deviations from this goal are entirely due to MTurk software and outside of our control.

Regardless of where your next preregistration is hosted, make it a good preregistration by including what should be included and excluding what should be excluded.



## Feedback

We would like to thank Stephen Lindsay and Simine Vazire for taking time out of their incredibly busy schedules to give us invaluable feedback on a previous version of this post.

## Subscribe to Blog via Email

Enter your email address to subscribe to this blog and receive notifications of new posts by email.

Subscribe

### Footnotes.

1. This is because conducting unplanned analyses necessarily inflates the probability that you will find a statistically significant relationship even if no relationship exists. [[↵](#)]
2. For good explanations of the virtues of preregistration, see Lindsay et al. (2016) <[.html](#)>, Moore (2016) <[htm](#)>, and van't Veer & Giner-Sorolla (2016) <[htm](#)>. [[↵](#)]
3. Contrary to popular belief, the job of your pre-registration is NOT to show that your predictions were confirmed. Indeed, the critical aspect of pre-registration is not the prediction that you register – many good preregistrations pose questions (e.g., “We are testing whether eating Funyuns cures cancer”) rather than hypotheses (e.g., “We hypothesize that eating Funyuns cures cancer”) – but the analysis that you specify. In hindsight, perhaps our preregistration website should have been called AsPlanned rather than AsPredicted, although AsPredicted sounds better. [[↵](#)]
4. Even complex studies should have a simple and clear preregistration, one that allows a reader to casually differentiate between confirmation and exploration. Additional complexities could potentially be captured in other secondary planning documents, but because these are far less likely to be read, they shouldn't obscure the core basics of the simple preregistration. [[↵](#)]
5. We recently updated the AsPredicted questions, and so this OSF template contains slightly different questions than the ones currently on AsPredicted. We advise readers who wish to use the OSF to answer the questions that are currently on <https://AsPredicted.org>. [[↵](#)]

## GET COLADA EMAIL ALERTS.

---

you@email.com

**Subscribe**

Join 5,153 other subscribers

## SOCIAL MEDIA

---

 We tweet new posts: [@DataColada](#)

 And mastopost'em: [@DataColada@mas.to](#)

 And link to them on our [Facebook page](#)

## RECENT POSTS

---

[108] MRAN is Dead, long live GRAN

---

[107] Meaningless Means #3: The Truth About Lies

---

[106] Meaningless Means #2: The Average Effect of Nudging in Academic Publications is 8.7%

---

[105] Meaningless Means #1: The Average Effect of Nudging Is  $d = .43$

---

[104] Meaningless Means: Some Fundamental Problems With Meta-Analytic Averages

GET BLOGPOST  
EMAIL ALERTS

your@email.com

**Submit**

Join 5,153 other subscribers

TWEETER &  
FACEBOOK

 We tweet new posts:

@DataColada



And link to them on our

Facebook page

POSTS ON SIMILAR  
TOPICS

### Preregistration

[101] Transparency Makes  
Research Evaluable: Evaluating  
a Field Experiment on Crime  
Published in *Nature*

---

[64] How To Properly  
Preregister A Study

---

[55] The file-drawer problem is  
unfixable, and that's OK

---

[44] AsPredicted: Pre-  
registration Made Easy

---

[12] Preregistration: Not just for  
the Empiro-zealots

SEARCH

Search ...

**Search**

© 2021, Uri Simonsohn, Leif Nelson, and Joseph Simmons. For permission to reprint individual blog posts on DataColada please contact us via email..