Running head: WILLINGNESS TO PRE-REGISTER

1

Will knowledge about more efficient study designs increase the willingness to pre-register?

Daniel Lakens<sup>1</sup>

<sup>1</sup> Eindhoven University of Technology

## Author Note

Funding for this research was provided by the Berkeley Initiative for Transparency in the Social Sciences and the Laura and John Arnold Foundation. All data and materials are available from https://osf.io/pwtrh/.

Correspondence concerning this article should be addressed to Daniel Lakens, IPO 1.24, Den Dolech 1, 5600 MB, Eindhoven, The Netherlands. E-mail: D.Lakens@tue.nl

### Abstract

Pre-registration is a straightforward way to make science more transparant, and control Type 1 error rates. Pre-registration is often presented as beneficial for science in general, but rarely as a practice that leads to immediate individual benefits for researchers. One benefit of pre-registered studies is that they allow for non-conventional research designs that are more efficient than conventional designs. For example, by performing one-tailed tests and sequential analyses researchers can perform well-powered studies much more efficiently. Here, I examine whether such non-conventional but more efficient designs are considered appropriate by editors under the pre-condition that the analysis plans are pre-registered, and if so, whether researchers are more willing to pre-register their analysis plan to take advantage of the efficiency benefits of non-conventional designs. Study 1 shows the large majority of editors judged one-tailed tests and sequential analyses to be appropriate in psychology, but only when such analyses are pre-registered. In Study 2 I asked experimental psychologists to indicate their attitude towards pre-registration. Half of these researchers first read about the acceptence of one-tailed tests and sequential analyses by editors, and the efficiency gains of using these procedures. However, learning about the efficiency benefits associated with one-tailed tests and sequential analyses did not substantially influence researchers' attitudes about benefits and costs of pre-registration, or their willingness to pre-register studies. The self-reported likelihood of pre-registering studies in the next two years, as well as the percentage of studies researchers planned to pre-register in the future, was surprisingly high. 47% of respondents already had experience pre-registering, and 94% of respondents indicating that they would consider pre-registering at least some of their research in the future. Given this already strong self-reported willingness to pre-register studies, pointing out immediate individual benefits seems unlikely to be a useful way to increase researchers' willingness to pre-register any further.

Keywords: Pre-registration, One-tailed Tests, Sequential Analyses, Meta-Science

Will knowledge about more efficient study designs increase the willingness to pre-register?

Pre-registration of hypotheses has been heralded as a promising solution to improve transparency in science (Chambers, Feredoes, Muthukumaraswamy, & Etchells, 2014; Nosek & Lakens, 2014). Clearly specifying the rules that will be used to terminate the data collection, and specifying the analysis plan before data is collected, can prevent undesirable research practices such as inflated of Type 1 errors and "hypothesizing after the results are known" (Kerr, 1998). Controlling alpha levels is beneficial for the scientific community in general, because widely used significance tests are difficult to interpret when the alpha level is not controlled. Although the benefits of pre-registration for a more reliable psychological science are clear, pre-registration is typically not advertised as a best-practice that leads to immediate individual benefits for researchers. A salient aspect of performing pre-registered studies is that researchers will have to learn where and how to pre-register their study, in addition to the time investment each pre-registration will take. However, there are also benefits to pre-registering your research. One of the biggest immediate rewards of pre-registered studies is that they allow non-conventional research designs that are more efficient than conventional designs. Researchers who pre-register their study design and analysis plan can profit from the immediate individual benefit of collecting data more efficiently by using sequential analyses and one-sided tests where useful and appropriate. In a time where researchers are increasingly required to perform high-powered studies to publish in top journals in psychology (Lindsay, 2015; Vazire, 2016), researchers can be expected to be motivated to perform high-powered studies efficiently.

Both sequential analyses (Lakens, 2014) and, whenever one has a directional prediction, one-sided tests (Knottnerus & Bouter, 2001; Maner, 2014; Rice & Gaines, 1994) can increase the efficiency with which researchers perform well-powered studies. In one-tailed tests, only predictions in one direction are tested (e.g., mean X will be higher than mean Y). In sequential analyses, researchers collect data, analyze it, and depending on the result, collect more data, while controlling their overall Type 1 error rate by using Bonferroni-like (but

slightly more efficient) adjustments to the alpha level at each look at the data. The efficiency benefit can be substantial: One-sided tests in sequential designs easily require 33% less participants when aiming for 80% power, and benefits increase when researchers aim for higher power or use sequential designs tailored to their research question. This is also the reason one-sided tests and sequential analyses are common in medicine, where the efficiency benefit has both a monetary component, but also an ethical component. As soon as there is evidence for a beneficial effect, all patients should receive the treatment, and the remaining patients should be enrolled in a new study that examines other possibly beneficial treatments (Jennison & Turnbull, 2000).

Best practice in medicine requires researchers to pre-register sequential analyses and one-tailed tests. It seems plausible (and this assumption is confirmed in Study 1) that editors at journals in psychology would only have a positive attitude toward the use of one-tailed tests and sequential analyses when these procedures are pre-registered. When these procedures are not pre-registered, they allow for additional ways to inflate the Type 1 error rate when analyzing data. If the time cost of performing pre-registered studies are an important reason why researchers choose not to pre-register their hypotheses, then explaining these immediate benefits of sequential analyses and one-tailed tests to researchers, while adding that these benefits are only likely to be accepted when pre-registered, might provide an incentive to pre-register their research, and facilitate the uptake of pre-registration in psychology.

I believe reducing the amount of time spent on data collection might be an immediate reward for many researchers, which could increase the willingness to pre-register studies. The hypothesis that more efficient study designs might increase the willingness to pre-register rests on several requirements. First, one-tailed tests and sequential designs should be considered acceptable, at least under some specified circumstances, in psychological science. Second, editors at psychology journals should have a positive attitude towards one-tailed tests and sequential designs only when these are pre-registered, and

indicate a negative attitude towards these more efficient designs when they are not pre-registered, thus raising the expectation that researchers can only yield the benefits of one-tailed tests and sequential designs when they pre-register their analysis plan. The assumptions of when and whether one-tailed tests are appropriate was examined in an open online discussion with interested researchers, and through a survey sent to editors at three top journals in psychology (Psychological Science, the Journal of Experimental Psychology: General, and Social Psychology and Personality Science).

A final assumption is that there is room for improvement. Researchers should not already pre-register most of their studies, and the likelihood that researchers will pre-register their analysis plan should not be at a ceiling. Data collected by Brett Buttliere and Jelte Wicherts<sup>1</sup> at the end of 2013 from 2302 researchers indicated 75% had never pre-registered a study. In a study conducted at the end of 2015 and early 2016, Hanne Watkins<sup>2</sup> found that the likelihood that researchers would pre-register a study in the next 6 months (from 1, very unlikely, to 7, very likely) was on average below the scale midpoint (M = 3.17, SD = 1.94). Based on these data, it seems there is still some room for improvement.

To examine whether pointing out the benefits of more efficient study designs would increase the willingness of researchers to pre-register their studies, in Study 2 a short online questionnaire was sent to corresponding authors of articles published in three top journals in psychology (the same three journals whose editors were approached in Study 1). Two versions of the questionnaire were created. In one, five questions were asked about whether researchers had ever pre-registered a study, the perceived costs and benefits of pre-registering an analysis plan, how likely they were to pre-register a study in the future, and an indication of the percentage of studies they would consider pre-registering. In the second condition, these questions were identical, but they were preceded by a short explanation and example of the benefits of one-sided tests and sequential designs, and a summary of the results of the

<sup>&</sup>lt;sup>1</sup>Buttliere, B., & Wicherts, J. M. (2015, June 17). Repligate survey. Retrieved from osf.io/sfkhe

<sup>&</sup>lt;sup>2</sup>Watkins, H. M. (2017, January 23). A Survey of Responses to the Replicability Debate in Psychology. Retrieved from osf.io/hjxkq

editor survey from Study 1. Previous research has shown that making cost-benefit analyses of using statistical approaches explicit can influence researchers' attitudes. For example, explicitly pointing out the problematic aspects of performing underpowered research changed researchers perception of job candidates (Gervais, Jewell, Najle, & Ng, 2015). Study 2 aimed to examine whether more efficient study designs would increase the willingness to pre-register studies, when these benefits were explicitly pointed out.

## Are One-tailed Tests and Sequential Analyses Appropriate in Psychology?

The discussion about whether one-tailed tests should be used in psychology has started more than 60 years ago (Burke, 1953; Eysenck, 1960; Goldfried, 1959; Jones, 1952, 1954; Kimmel, 1957; Marks, 1953), and remains a topic of debate until today. It seems likely that the researchers who choose to dedicate their time to publish on this topic are those that have the most extreme opinions, but relatively little is known about the average researcher's attitudes towards one-tailed tests. We can see one-tailed tests are rarely used in psychology, maybe because researchers prefer to avoid this discussion during the review process. In medicine, one-tailed tests are used more commonly, especially in large randomized controlled Phase III trials. Greenland, Senn, Rothman, Carlin, Poole, Goodman, and Altman (2016) respond to the statement that "One should always use two-sided p-values" with a resounding "No!", indicating that when a hypothesis is directional, a one-tailed test is required<sup>3</sup>.

Since one-sided tests are only appropriate whenever null-hypothesis significance tests are appropriate, we have to assume that a null-hypothesis test answers a question a researcher is interested in, and that the null-hypothesis is both plausible and interesting. Beyond this, both in the scientific literature, as in our online discussion, there was general agreement about the need to pre-register one-sided tests. For example, Marks (1953) notes: "It must be emphasized that the one-tailed test is not justified unless the prediction is made prior to the data'. Similarly, Dubey (1991) recommends: "The choice of the test (one-sided

<sup>&</sup>lt;sup>3</sup>Although not all authors of the publication agree (Senn, personal communication, 1-1-2017: https://twitter.com/stephensenn/status/815583465240494080)

or two-sided) should be specified in the protocol, thereby eliminating any possibilities of post hoc manipulations."

I will briefly mention some general points that researchers considered in our online discussion about whether one-tailed tests are appropriate<sup>4</sup>. One-tailed tests have some clear benefits. They require a smaller sample size than a two-tailed test to achieve the same power, while keeping the error rate constant (e.g., at 5%). This means that one-tailed tests will often be an optimal choice from an Neyman-Pearson perspective, and this efficiency has an ethical component when funding comes from tax money. Another benefit of one-tailed tests is that they are logically consistent with the conclusion we want to draw from studies. Researchers typically conclude one mean is larger than another after observing a significant result, while the logically correct conclusion after a significant result in a two-tailed test is that the means differ from zero, but the direction in which they differ has not been tested. One-tailed tests also have the statistically attractive property that they have a direct Bayesian interpretation.

But there are also downsides. A smaller sample size means less precise data (i.e., wider confidence intervals), and less evidence. Assuming normally distributed data, the Z value required to reach statistical significance is Z=1.96 for a two-tailed tests, but only 1.645 for a one-tailed test. This means that, even though the Type 1 error rate is the same for one-tailed and two-tailed tests, a lower evidence threshold is passed. This bothers neo-Fisherians, who interpret p-values as evidence (e.g., see Lombardi & Hurlbert, 2009). It also means it is slightly easier to find a Type 1 error in the expected direction (e.g., one that confirms one's beliefs). Another problem is whether effects in the opposite direction can be described and reported. The difference between hypothesis testing and estimation is relevant here - an unexpected effect might not confirm a hypothesis, but the descriptives related to the effect size might still be interesting. A final downside of one-tailed tests is that they signal interestedness of a researcher, when disinterestedness might be more desirable.

<sup>&</sup>lt;sup>4</sup>See http://bit.ly/2lx8r9x. I'd like to thank Ian Hussey, Marc Halusic, Derek Berger, Mark Kelson, Gjalt-Jorn Peters, L. J. Zigerell, Rickard Carlsson, and Marcel van Assen for their contributions.

There is almost no work on the acceptance of sequential analyses in psychology. Sequential analyses from a frequentist and Bayesian perspective have been recommended for efficiency gains (Lakens, 2014; Schonbrodt, Wagenmakers, Zehetleitner, & Perugini, 2015), and have been recommended in at least one editorial in psychology (Giner-Sorolla, 2016). In essence, controlling alpha levels in sequential analyses is exactly the same as controlling for multiple comparisons as a result of comparing more than 2 conditions, or analyzing multiple dependent variables. However, since they lead to lower sample sizes (on average) than non-sequential designs, one can again consider a difference in attitudes between researchers who use a Neyman-Pearson error control approach to statistical inferences, and statistical approaches that more strongly value the amount of evidence in the data. However, the amount of evidence one desires and the error rate which is deemed acceptable are too different goals, and require a different approach when designing a study.

It is perhaps surprising that no one has ever simply asked researchers whether one-tailed tests and sequential analyses are deemed appropriate. Even though statistics is not necessarily a democratic enterprise, science relies on peer review, and thus it is arguably of interest to researchers to know what their peers, and especially those working in editorial boards of top journals in their field, think about how appropriate one-tailed tests are. Furthermore, recent developments, such as the ability to pre-register a study, might have influenced these attitudes. Therefore, in Study 1 I asked members of the editorial board of three top journals in psychology to answer seven short questions about the appropriateness of one-tailed tests and sequential analyses.

## Study 1: Editor Survey

Participants and Procedure. All names and e-mail addresses of members listed as members of the complete editorial boards of three top journals in psychology (Psychological Science, the Journal of Experimental Psychology: General, and Social Psychology and Personality Science) were manually collected. All editors received an e-mail

on the 21st of February, 2017, and a reminder on the 28th of February, 2017. Respondents were informed that for each completed questionnaire \$5 would be donated to a charity that aims to advance science education. All data received before the 7th of March, 10:00 am GMT+1 was included in the analysis. The questionnaire was administered through SurveyMonkey. Overall, 327 emails were sent, of which 235 were opened in an e-mail client, and 87 editors completed the survey of seven questions. All anonymized data and materials are available from https://osf.io/pwtrh/.

Results. The first question asked whether one-tailed tests were ever appropriate in psychology. There was no perfect agreement, with 80 editors indicating one-tailed tests were appropriate, and 6 editors who were of the opinion that one-tailed tests are never appropriate in psychology (1 editor indicated not to know whether one-tailed tests were ever appropriate). It is interesting to see that there is substantial agreement on this issue, as opposed to the often polarized debate in the literature on this topic. Even though around 92% of editors accept one-tailed tests, some editors think the use of one-tailed tests is never appropriate. Editorial policies do not explicitly state whether and when one-tailed tests are considered appropriate, but the small diversity in answers suggests it might be useful to specify a general policy on the journal level. The mean attitude towards one-tailed tests of editors when one-tailed tests were pre-registered was more positive (M = 5.17, SD = 1.56)than when one-tailed tests were not pre-registered (M = 2.70, SD = 1.43, r = 0.35). The difference between measurements (M = 2.47, SD = 1.71) was large, t(86) = 13.47, p < 0.001, Hedges'  $g_z = 1.43$ , 95% CI [1.13;1.73], see Figure 1. Notably, although some editors indicated pre-registration did not change their attitude towards one-tailed tests (n = 15), all other editors (n = 72) indicated pre-registration led to a more positive attitude about one-tailed tests. The size of the effect of pre-registration is substantial, with the general attitude being negative without pre-registration, but positive with pre-registration. A final question in the survey concerned in which percentage of studies where t-tests were reported, editors felt that, on average, one-tailed tests were appropriate. We see quite some variation in the responses

(see Figure 2), with many editors believing one-tailed tests are rarely appropriate (these include the editors who think one-tailed tests are never appropriate), but with answers ranging across the spectrum (M = 33.30, median = 20).

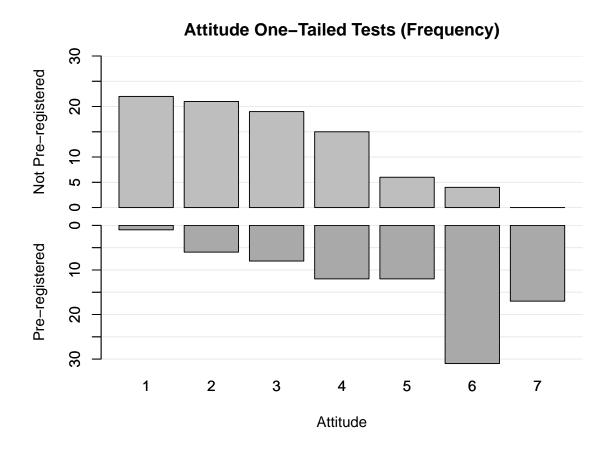


Figure 1. Attitude one a scale from 1 (very negative) to 7 (very positive) about reporting one-tailed tests when the analysis plan was not pre-registered (top) or pre-registered (bottom).

With respect to attitudes towards sequential analyses, only 2 editors indicated sequential analyses were never appropriate in psychology, 13 editors indicated they did not know whether sequential analyses were appropriate, and 71 editors indicated sequential analyses are appropriate. The slightly larger number of editors who were unsure about sequential analyses is not surprising, given that sequential analyses are not yet widely used in psychology. As part of the current project, I have created educational materials that explain the general idea behind sequential analyses and one-tailed tests that are now part of

# How often are one-tailed tests appropriate?

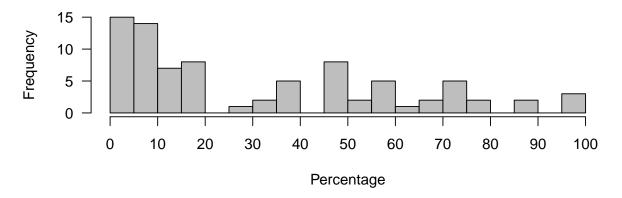


Figure 2. Percentage of studies in which t-tests were reported, and where editors considered one-tailed tests to be appropriate.

a Coursera MOOC "Improving Your Statistical Inferences". Researchers interested in an introduction to sequential analyses can watch videos 3.1 and 5.2, and complete assignment  $3.2^5$ .

The mean attitude towards sequential analyses (on a scale from 1, "strongly negative" to 7, "strongly positive") was quite negative when this procedure was not pre-registered (M = 2.85, SD = 1.41) but was quite positive when pre-registered (M = 5.20, SD = 1.29, r = 0.22). The difference between measurements (M = 2.35, SD = 1.69) was analyzed with a dependent t-test, t(85) = 12.87, p < 0.001, Hedges'  $g_z = 1.38$ , 95% CI [1.08;1.67].

Although again some researchers indicated pre-registration did not change their attitude towards sequential analyses (n = 16), all other researchers who answered this question (n = 70) indicated pre-registration led to a more positive attitude about reported sequential analyses (see Figure 3).

The results of the questionnaire are clear: One-tailed tests and sequential analyses are considered appropriate by the majority of the editors. Attitudes towards these procedures

<sup>&</sup>lt;sup>5</sup>The course is available, for free, at https://www.coursera.org/learn/statistical-inferences

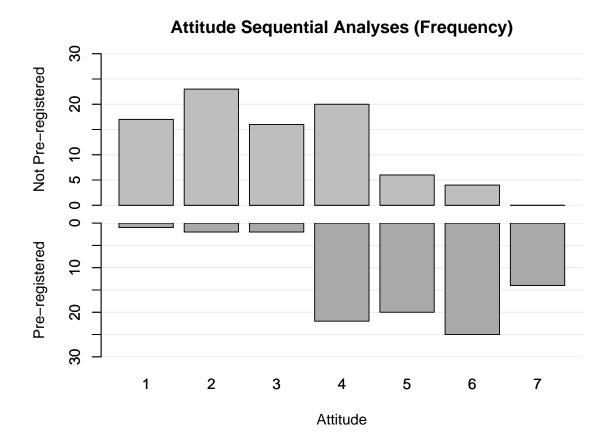


Figure 3. Attitude one a scale from 1 (very negative) to 7 (very positive) about reporting sequential analyses when the analysis plan was not pre-registered (top) or pre-registered (bottom).

were greatly dependent upon whether the analysis plan was pre-registered or not, with attitudes being quite negative without pre-registration, and quite positive with pre-registration. This suggests that pre-registration might be considered a pre-condition of publishing articles in which one-tailed tests or sequential analyses are reported. As a consequence, the efficiency benefit that one-tailed tests and sequential analyses provide might be a strong incentive for researchers to pre-register. In Study 2, I examine whether learning about the efficiency benefits of one-tailed tests and sequential analyses, and the acceptance of these procedures by editors when the analyses are pre-registered, might increase the willingness of researchers to pre-register their studies.

### Study 2: Researchers

References for the last 1000 published articles in Psychological Participants. Science and JEP:General, and all (662) articles in Social Psychological and Personality Psychology, including corresponding authors and e-mail addresses, were downloaded from WebOfScience. All names of first authors and e-mail addresses of corresponding authors were automatically retrieved, and manually checked. Because the goal was to send personalized email invitations, combinations where the first author's name did not match the e-mail address of the corresponding author were manually deleted, and double entries were removed. All first authors received an e-mail on the 28st of February, 2017, and a reminder on the 7th of March, 2017. All data received before the 14th of March, 10:00 am GMT+1 was included in the analysis. The questionnaire was administered through SurveyMonkey. Overall, 1802 emails were sent, of which 993 were opened in an e-mail client (the low percentage might in part be due to changes in e-mail addresses over time), and 409 researchers completed the survey of 5 questions. A total of 6 responses were removed because researchers either did not answer, or answered both "yes" and "no" to the first question asking if they had ever performed a pre-registered study, leaving 184 observations in the control condition, and 219 in the experimental condition. Respondents were informed that for each completed questionnaire \$2 would be donated to a charity that aims to advance science education.

**Procedure.** After an introduction to the questionnaire, and a brief explanation what pre-registration entails, half of the researchers immediately started with the questionnaire (the control condition). For the other half of the researchers, the benefit of one-tailed tests and sequential analyses was explained through a numerical example of the average sample size needed for a traditional two-tailed non-sequential design, and a one-tailed sequential design. Subsequently, a summary of Study 1 was presented (based on the completed questionnaires after 1 week, which was basically identical to the final results reported above), with the goal to communicate that these practices were deemed appropriate by editors, but only when pre-registered. Subsequently, researchers in the experimental condition were asked

to answer the same five questions as researchers in the control condition. All anonymized data and materials are available from https://osf.io/pwtrh/.

Results. Of primary interest was whether researchers in the experimental conditions, who learned about the efficiency benefits of one-tailed tests and sequential designs, would judge pre-registering their research as more beneficial, indicate they were more likely to pre-register a study in the future, or plan to perform a greater percentage of pre-registered studies in the future. I was especially interested in whether researchers who had never performed a pre-registered experiment in their own lines of research would show a greater willingness to pre-register.

There were 223 researchers who had never pre-registered, 59 who had pre-registered a study as part of a collaborative research project, 72 who had pre-registered a study in their own research line, and 49 who had both pre-registered a study in their own research line, and participated in a collaborative research project that was pre-registered. Focusing on those researchers who had never pre-registered a study in their own research line, I examined whether hearing about the benefits of one-tailed tests and sequential analyses, and the positive attitude of editors towards these approaches as long as they were pre-registered, influenced the answers researchers gave regarding the benefits and costs of pre-registration, how likely they were to pre-register, and the percentage of studies they planned to pre-register in the future. None of the answers to the four questions revealed statistically significant differences between the experimental and control condition (see Table 1).

Equivalence tests were performed using the two-one-sided test (TOST) approach, with an upper equivalence bound of Cohen's d=0.3. The smallest effect size of interest is a subjective judgment, and especially difficult to determine when both the population standard deviation and the total sample size is unknown (Lakens, 2017). Assuming a standard deviation between 1 and 1.5 (not uncommon for a 5 or 7 point scale) the effect size of 0.3 would translate into a difference of 0.3 to 0.45 on an answer scale. We would need approximately 191 respondents in each condition (or 382 in total) to have 80% power to

Table 1

Means, standard deviations, and sample sizes for answers to the four survey questions as a function of condition

Question	Condition	Mean	SD	n	t	p	g	95% CI
Benefits	Control	3.45	1.13	150	0.74	0.77	-0.08	-0.3;0.14
	Experimental	3.36	1.08	180				
Costs	Control	2.60	1.05	148	0.42	0.34	0.05	-0.17;0.27
	Experimental	2.56	0.89	180				
Likely to Preregister	Control	5.28	1.72	150	1.83	0.97	-0.20	-0.42;0.02
	Experimental	4.92	1.83	181				
Percentage Preregister	Control	46.85	30.56	143	0.23	0.59	-0.03	-0.24;0.19
	Experimental	46.07	30.71	180				

reject effects of this or greater magnitude.

Four Welch's independent t-tests using the TOST approach indicated statistical equivalence for the difference between conditions for the perceived benefits, t(311.65) = -3.45, p < 0.001, costs, t(288.61) = -2.26, p = 0.012, how likely researchers were to pre-register, t(323.68) = -4.55, p < 0.001, and percentage of studies they planned to pre-register, t(305.34) = -2.90, p = 0.002. We can reject effects large enough to matter in practice, and conclude there is no meaningful effect.

Spearman's correlations between the answers to the four questions shows that the likelihood researchers will pre-register in the future, and the percentage of studies the plan to pre-register, are strongly monotonicly related with the perceived benefits of pre-registration (r = 0.57, 95% CI [0.49;0.66], and r = 0.54, 95% CI [0.46;0.62], respectively), and somewhat less strongly related to the costs (r = -0.26, 95% CI [-0.33;-0.17], and r = -0.28, 95% CI [-0.37;-0.20], respectively). Perceived benefits and costs were negatively related, as expected,

but not strongly so (r = -0.27, 95% CI [-0.35; -0.18).

When exploring difference in responses between previous experience with pre-registration, we see a clear trend where reasearchers who have pre-registered studies in their own research indicate pre-registration is more beneficial, and indicate higher a higher likelihood of pre-registering studies in the future, and higher percentage of studies for which they would consider pre-registering (see Table 2). This shows the a-priori idea to focus the main analysis on people who has not yet pre-registered studies in their own lines of research was valid.

An interesting observation is that researchers report to be quite likely to pre-register a study in the next two years (M = 5.34, SD = 1.76). Even though this percentage must be interpreted in light of possible self-selection, this is a much higher estimate than in the data collected by Watkins (M = 3.17, SD = 1.94), who asked about planned pre-registrations in the next 6 months. Perhaps attitudes have changed over the last 12 months, or researchers think it is more likely that pre-registration will be useful for some future study, even when it is not considered very beneficial for the studies they have planned in the next 6 months. Whether people are really more likely to pre-register a study within a two-year time period, compared to a 6 month period, or are simply displaying temporal discounting, would be an interesting topic of research.

In the data collected by Buttliere and Wicherts at the end of 2013 only 25% of researchers had experience pre-registering their research. In the current data collected in early 2017, 45% of the researchers had experience pre-registering their research, which is a substantial increase in just over three years. Although we have to interpret these numbers in light of possible self-selection, both pro-ponents as opponents of pre-registration might have been motivated to participate. Only 15 researchers indicated they would not consider to

Table 2

Means, standard deviations, and sample sizes for answers to the four survey questions as a function of previous experience with pre-registration

Question	Pre-registration Experience	Mean	SD	N
Benefits	No Experience		1.12	222
	Collaborative Project	3.15	1.00	59
	Own Research	3.93	0.98	72
	Own Research and Collaborative Project	4.16	0.83	49
Costs	No Experience	2.65	0.98	220
	Collaborative Project	2.64	0.89	59
	Own Research	2.32	0.80	72
	Own Research and Collaborative Project	2.16	0.87	49
Likely to Pre-register	No Experience		1.81	223
	Collaborative Project	5.49	1.33	59
	Own Research	6.54	0.90	72
	Own Research and Collaborative Project	6.82	0.63	49
Percent Pre-register	No Experience	40.68	29.50	215
	Collaborative Project	44.95	27.15	59
	Own Research	61.31	24.00	72
	Own Research and Collaborative Project	73.33	24.97	49

pre-register any of their studies in the future<sup>6</sup>, meaning that 94% of researchers would consider pre-registering at least some of their research. The median response to the question which percentage of studies they would consider pre-registering in the future was 50% (see lower right panel of Figure 4). This suggests a surprisingly high self-reported willingness to pre-register studies, especially if we take into account that there is no a-priori reason to assume a pre-registration rate of 100% is necessary. Purely descriptive studies where all data is reported might not require pre-registration, small pilot studies to test the feasibility of a paradigm are not designed to test a hypothesis, and some researchers might only publish studies when they have directly replicated every test in an independent sample. If this high self-reported willingness to consider pre-registration of analysis plans materializes, this can be considered a substantial change in the way psychological research is done, and there might not be a strong need to further increase the willingness to pre-register studies. On the other hand, when we observe an attitude-behvior gap in the future, it may be that immediate efficiency benefits achieved by pre-registering studies would increase pre-registration behavior, even when it did not influence attitudes in this study.

#### General Discussion

Based on the current data, it seems unlikely that the efficiency benefit associated with one-tailed tests and sequential analyses are an efficient way to improve the willingness to pre-register studies. There are several reasons why this might be the case. First of all, data collection might not be too costly for many researchers, making efficiency benefits trivial. Indeed, if data collection is cheap (e.g., when you perform online studies where participants are paid very little) the difference between 172 participants and 230 participants might not be something you care about. As an illustration, one researcher contacted me to say that it was more efficient to simply replicate their own experiment, than to pre-register the study.

<sup>&</sup>lt;sup>6</sup>An additional 8 researchers skipped this question. Because the default slider position was at 0 but not moving the slider was recorded as a non-response in SurveyMonkey, these researchers might also have desired to answer 0%.

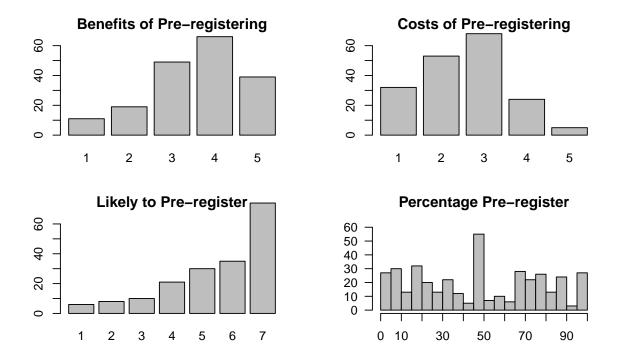


Figure 4. Frequencies of responses for the four questions about pre-registration.

Another possible reason for the lack of an effect of pointing out the efficiency benefits of pre-registration might be that the perceived benefits of pre-registration are less material, and more idealistic, for most researchers. If you value transparency in science, pre-registration is a valuable tool, regardless of the time and money it could potentially save. As indicated in the grant proposal for this project, I believe that despite the fact that the hypothesis was falsified by the data, the outcome of the project is relevant. The current data at least tentatively suggest immediate individual benefits are not the best way to convince people to pre-register their research, and alternative options (e.g., referring to shared norms) might be a more beneficial approach.

Researchers indicated it was rather likely that they would pre-register a study in the next two years, to a substantially greater extent than observed by Watkins (2017) when asked about their plans in the next six months. Whether these self-reported expectations really materialize remain to be seen. Another possibility is that the difference between how likely researchers think they are to pre-register a study in the next 2 years, versus the next

six months, is an example of temporal discounting (Frederick, Loewenstein, & O'Donoghue, 2002). A follow-up study a few years from now, where the current respondents are asked which percentage of studies they actually pre-registered, could provide an answer to this question.

Although we contacted a representative sample of editorial board members and researchers in studies 1 and 2, one limitation of this research is the possibility that self-selection played a role in the people who chose to complete the questionnaire. In the editor survey, only one editor indicated not knowing whether one-tailed tests were appropriate in psychology, which is a rather low number compared to the extensive discussion, and sometimes strong disagreement, in the scientific literature. Since the invitation to the questionnaire clearly explained the topic was about one-tailed tests and sequential analyses it is possible only editors interested in, and knowledgeable about, this topic filled in the questionnaire. We might expect the inclusion of more individuals without clear preferences to reduce the observed effect sizes somewhat, although given the very large effects observed, this does not seem too problematic.

In the researcher survey, there is the possibility that the experimental manipulation might have been most effective on researchers without strong opinions about pre-registration, who might have been less interested in filling out this survey. Self-selection might have positively biased answers about the perceived benefits and likelihood researchers would pre-register in the future. Furthermore, editors and researchers from only three journals were sampled (even though both Psychological Science and JEP:General publish on a wide range of topics in psychology). Although the contacted editors are often in editorial boards of other psychology journals, and the same researchers publish in other psychology journals, the selection of journals limits the generalizability of these results.

One-tailed tests and sequential analyses might not increase researchers' willingness to preregister their hypotheses, they still yield an important efficiency benefit. The ethical arguments to spend research money as efficiently as possible that have been brought to bear in medicine (Jennison & Turnbull, 2000; Knottnerus & Bouter, 2001), also apply to psychology (Lakens, 2014). The results from Study 1 suggest that one-tailed tests and sequential analyses are deemed appropriate by most editors in psychology, and whenever researchers have the goal to determine whether the null-hypothesis can be rejected with a certain maximum error rate, their sample size justification should include a careful cost-benefit analysis, where the consideration of one-tailed tests and/or sequential analyses might be warranted.

Even though one-tailed tests and sequential analyses were deemed appropriate by most editors contacted in the current survey (as long as these analyses were pre-registered), communicating this efficiency benefit to researchers was not substantial enough to influence their willingness to pre-register. It seems most fruitful to explore other avenues to stimulate the adoption of pre-registration in psychology. One possibility is to simply require pre-registration, as is the case in other disciplines such as medicine. Although it remains to be seen whether the high self-reported willingness to pre-register will materialize in the future, there seems to a positive trend towards a greater willingness to pre-register studies among psychologists.

### References

- Burke, C. J. (1953). A brief note on one-tailed tests. Psychological Bulletin, 50(5), 384-387.
- Chambers, C. D., Feredoes, E., Muthukumaraswamy, S. D., & Etchells, P. (2014). Instead of playing the game it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. AIMS Neuroscience, 1(1), 4-17.
- Eysenck, H. J. (1960). The concept of statistical significance and the controversy about one-tailed tests. Psychological Review, 67(4), 269.
- Frederick, S., Loewenstein, G., & O'donoghue, T. (2002). Time discounting and time preference: A critical review. Journal of Economic Literature, 40(2), 351-401.
- Gervais, W. M., Jewell, J. A., Najle, M. B., & Ng, B. K. L. (2015). A Powerful Nudge?

  Presenting Calculable Consequences of Underpowered Research Shifts Incentives Toward

  Adequately Powered Designs. Social Psychological and Personality Science, 6(7), 847-854.

  https://doi.org/10.1177/1948550615584199
- Giner-Sorolla, R. (2016). Approaching a fair deal for significance and other concerns. Journal of Experimental Social Psychology, 65, 1-6. https://doi.org/10.1016/j.jesp.2016.01.010
- Goldfried, M. R. (1959). One-tailed tests and "unexpected" results. Psychological Review, 66(1), 79-80. https://doi.org/10.1037/h0038521
- Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. European Journal of Epidemiology, 31(4), 337-350. https://doi.org/10.1007/s10654-016-0149-3
- Jennison, C., & Turnbull, B. W. (2000). Group sequential methods with applications to clinical trials. Boca Raton: Chapman & Hall/CRC. Jones, L. V. (1952). Test of

- hypotheses: one-sided vs. two-sided alternatives. Psychological Bulletin, 49(1), 43-46. https://doi.org/http://dx.doi.org/10.1037/h0056832
- Jones, L. V. (1954). A rejoinder on one-tailed tests. Psychological Bulletin, 5(6), 585-586.
- 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
- Kimmel, H. D. (1957). Three criteria for the use of one-tailed tests. Psychological Bulletin, 54(4), 351-353. https://doi.org/10.1037/h0046737
- Knottnerus, J. A., & Bouter, L. M. (2001). The ethics of sample size:: Two-sided testing and one-sided thinking. Journal of Clinical Epidemiology, 54(2), 109-110.
- Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses: Sequential analyses. European Journal of Social Psychology, 44(7), 701-710. https://doi.org/10.1002/ejsp.2023
- Lakens, D. (2017). Equivalence tests: A practical primer for t-tests, correlations, and meta-analyses. Social Psychological and Personality Science. DOI: 10.1177/1948550617697177
- Lindsay, D. S. (2015). Replication in Psychological Science. Psychological Science, 26(12), 1827-1832. https://doi.org/10.1177/0956797615616374
- Lombardi, C. M., & Hurlbert, S. H. (2009). Misprescription and misuse of one-tailed tests. Austral Ecology. https://doi.org/10.1111/j.1442-9993.2009.01946.x
- Maner, J. K. (2014). Let's Put Our Money Where Our Mouth Is: If Authors Are to Change Their Ways, Reviewers (and Editors) Must Change With Them. Perspectives on Psychological Science, 9(3), 343-351. https://doi.org/10.1177/1745691614528215
- Marks, M. R. (1953). One- and two-tailed tests. Psychological Review, 60(3), 207-208. https://doi.org/10.1037/h0054653

- 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
- Rice, W. R., & Gaines, S. D. (1994). "Heads I win, tails you lose": testing directional alternative hypotheses in ecological and evolutionary research. Trends in Ecology & Evolution, 9(6), 235-237. https://doi.org/10.1016/0169-5347(94)90258-5
- Schonbrodt, F. D., Wagenmakers, E. J., Zehetleitner, M., & Perugini, M. (2015). Sequential Hypothesis Testing With Bayes Factors: Efficiently Testing Mean Differences.

  Psychological Methods. Retrieved from <a href="http://www.ncbi.nlm.nih.gov/pubmed/26651986">http://www.ncbi.nlm.nih.gov/pubmed/26651986</a>
- Vazire, S. (2016). Editorial. Social Psychological and Personality Science, 7(1), 3-7. https://doi.org/10.1177/1948550615603955