Positive Results in Registered Reports

Anne Scheel[1], Mitchell Schijen[1], & Daniël Lakens[1]

[1] Eindhoven University of Technology

Author Note

Correspondence concerning this article should be addressed to Anne Scheel, Den Dolech 1, Atlas 9.417, 5600 MB, Eindhoven, The Netherlands. E-mail: a.m.scheel@tue.nl

Abstract

XXXXXXXXXXXXXXXXXXX

Keywords: keyword 1, keyword 2, keyword 3

Word count: XXX

Positive Results in Registered Reports

Analysis

We planned to test our hypothesis in the following way (quoting directly from our preregistration, https://osf.io/sy927, point 9):

> A one-sided proportion test with an alpha level of 5% will be performed to test whether the positive result rate (full or partial support) of Registered Reports in psychology is statistically lower than the positive result rate of conventional reports in psychology. In addition to testing if there is a statistically significant difference between RRs and conventional reports, we will test if the difference is smaller than our smallest effect size of interest using an equivalence test for proportion tests with an alpha level of 5% (Lakens, Scheel, & Isager, 2018). We determined our smallest effect size of interest to be the difference between the positive result rate in psychology (91.5%) and the positive result rate in general social sciences (85.5%) as reported by Fanelli (2010), i.e. a difference of $91.5\% - 85.5\% = 6\%$. The rationale for choosing general social sciences as a comparison is that this discipline had the lowest positive result rate amongst the 'soft' sciences (Fanelli, 2010). The exact percentage for general social sciences was extracted from Figure 1 in Fanelli (2010) using the software WebPlotDigitizer (Rohatgi, 2018).

We would accept our hypothesis that RRs have a lower positive result rate than SRs if we found a negative difference between RRs and SRs that was significantly different from 0 *and* not statistically equivalent to a range from -6% to $+6\%$ (both at $\alpha = 5\%$).

Results

Confirmatory Analyses

146 out of 152 SRs and 31 out of 71 RRs had positive results, meaning that the positive result rate was 96.05% for SRs (95% CI [91.61, 98.54]) and 43.66% for RRs (95% CI [31.91, 55.95]; see Fig. 1). The preregistered one-sided proportions test with an alpha level of 5% showed that this difference of −52.39% was statistically significant, $\chi^2 = 77.96$, $p < .001$. Unsurprisingly, the difference was not statistically equivalent to a range between −6% and 6% at $\alpha = 5\%$, $z = 7.61$, $p > .999$, meaning that we cannot reject differences more extreme than 6%. We thus accept our hypothesis that the positive result rate in RRs is lower than in SRs.

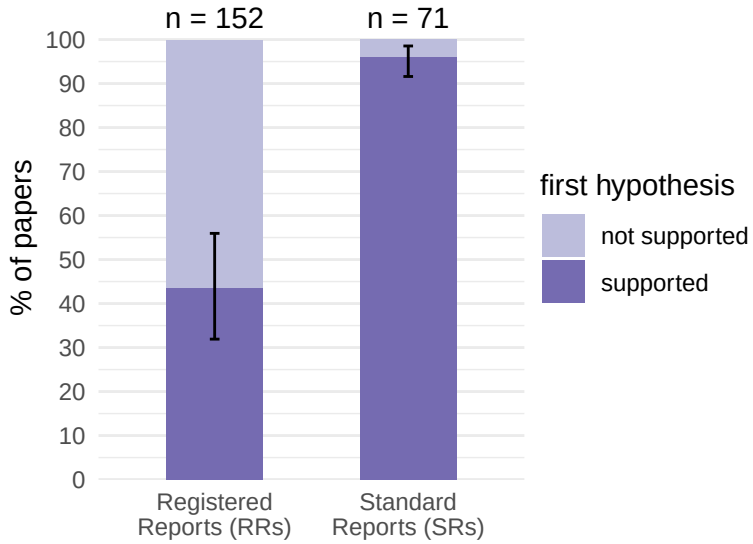


Figure 1. Positive result rates for standard reports and Registered Reports. Error bars indicate 95% confidence intervals around the observed positive result rate.

Exploratory Analyses

As expected, direct replications were much more common among RRs than SRs: 41 out of 71 RRs (57.75%), but only 4 out of 152 SRs (2.63%) were classified as direct

Table 1
Positive results in original studies vs replication studies

	original studies				replication studies			
	n	supported	%	95% CI	n	supported	%	95% CI
SRs	148	142	95.95	91.39; 98.50	4	4	100.00	39.76; 100.00
RRs	30	15	50.00	31.30; 68.70	41	16	39.02	24.20; 55.50

Note. SRs = standard reports, RRs = Registered Reports

replications of previously published work. However, this difference cannot account for the stark overall difference between SRs and RRs described above: Although replication RRs in our sample indeed had a lower positive result rate than original RRs (see Table 1), the difference between original SRs and original RRs – 45.95% – was still significantly different from 0 ($\chi^2 = 46.28$, $p < .001$) and not statistically equivalent to a range between -6% and 6% ($z = 4.31$, $p > .999$), both at $\alpha = 5\%$.

Since our SR sample represents a direct replication of Fanelli (2010) for the discipline Psychiatry & Psychology, another interesting question to ask is how our results compare to Fanelli's. The difference between the positive result rate for SRs in our sample (96.05%) and Fanelli's (91.49%) is 4.56%. This difference is not significantly different from 0 in a two-sided proportions test ($\chi^2 = 1.91$, $p = .167$) but also not statistically equivalent to a range between -6% and 6% ($z = -3.73$, $p = .306$), both at $\alpha = 5\%$. In other words, we can neither reject the hypothesis that the positive result rates of the two populations are the same, nor that there is a difference of at least $\pm 6\%$ between them. The data are inconclusive.

Finally, we analysed the language that was used to introduce or signal hypotheses in RRs. Does Fanelli's search phrase 'test* the hypothes*' capture hypothesis-testing RRs reasonably well? The answer is a resounding 'no': Searching the abstracts, titles, and keywords of the RR sample showed that only 2/71 RRs would have been detected with this search phrase. Which analogous hypothesis-introduction phrases are used in RRs, and can they be summarised in a convenient way? To get an overview, we stripped the hypothesis

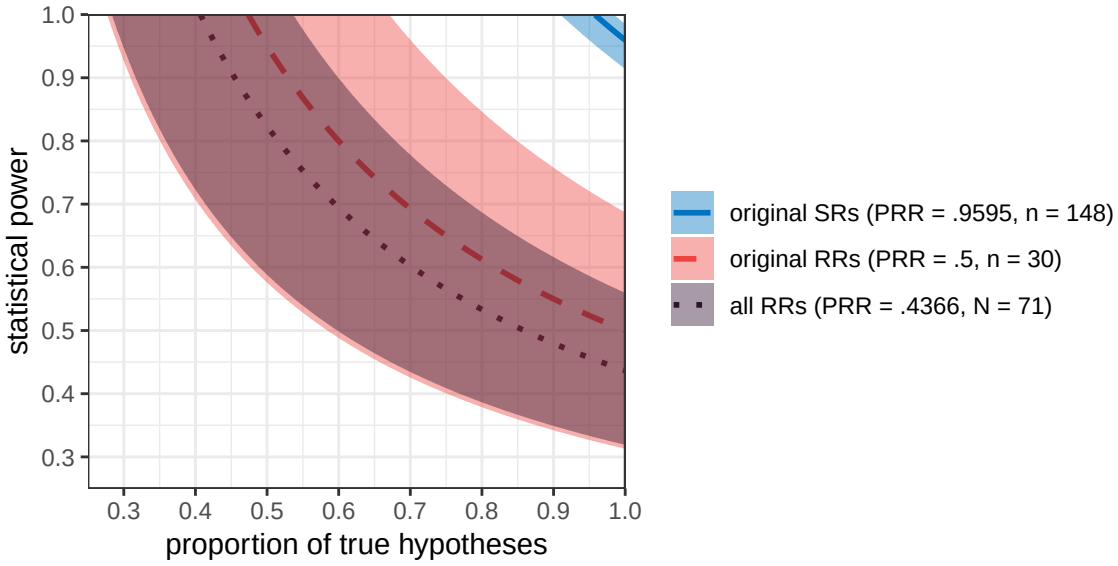
quotes of RRs from all content-specific information and extracted 'minimal' phrases that most distinctively indicated that a hypothesis was being described. For example, from the hypothesis quote '(f)or Study 1, we predicted that participants reading about academic (vs. social) behaviors would show a better anagram performance' we extracted the hypothesis-introduction phrase 'predicted that'.

For the majority of RRs (49), we identified one hypothesis-introduction phrase; the remaining ones used two (16 RRs), three (4 RRs), or four (1 RR) different phrases or had no identifiable hypothesis introduction (1 RR). In this total set of 97 hypothesis introductions, we found 64 unique phrases, showing substantial linguistic variation (see Tables 2 and 3). In order to condense the information further, we listed all unique word stems (e.g., the word stem 'hypothes*' captures the words 'hypothesis', 'hypotheses', 'hypothesize', 'hypothesized', and so on) and analysed their frequency among all hypothesis introductions. Excluding words that are common but too unspecific by themselves, such as 'that', 'to', or 'whether', the five most frequent word stems were 'hypothes*' (34 occurrences), 'replicat*' (24), 'test*' (20), 'examine*' (8), and 'predict*' (8). Clearly 'test*' and 'hypothes*' are quite popular, yet they co-occurred only 8 times and more than half of all hypothesis introductions (51/97) contained neither word. Interestingly, the frequency of these two words differed between original studies and direct replications: 30 out of 43 (69.77%) hypothesis introductions found in original RRs contained either 'test*' or 'hypothes*' or both, while only 16 out of 54 (29.63%) hypothesis introductions in direct replication RRs did.

We noticed that direct replication RRs generally tended to use different language to describe their hypothesis (or rather their goal). As the high frequency of the word stem 'replicat*' suggests, these studies were often not framed as new tests of a previously tested hypothesis, but as attempts to replicate a previously documented effect or to repeat a previously conducted experiment. Tables 2 and 3 list all unique hypothesis introductions and their frequency for original RRs and direct replication RRs, respectively, grouped by the five

most frequent word stems ('hypothes*', 'replicat*', 'test*', 'examine*', 'predict*'). Using five as a cutoff value is an arbitrary decision, but we believe that it strikes a reasonable balance between condensing the information and doing the variance of the data justice.

Discussion



References

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS ONE*, *5*(4), e10068. doi:10.1371/journal.pone.0010068

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. doi:10.1016/j.ympev.2015.01.015

Rohatgi, A. (2018). *WebPlotDigitizer - Web Based Plot Digitizer*. Austin, Texas, USA. Retrieved from https://automeris.io/WebPlotDigitizer

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS ONE*,

Table 2
Hypothesis introduction phrases in original Registered Reports (testing new hypotheses)

		source		
core word(s)	introduction phrase	abstract	full text	total
hypothes*, test*		3	2	5
	test of ... hypotheses	0	1	1
	test of ... hypothesis	1	0	1
	test the hypothesis that	1	0	1
	tested ... hypotheses	0	1	1
	tested the hypothesis that	1	0	1
hypothes*		5	12	17
	(Hypothesis 1)	0	1	1
	Hypothesis 1 (H1):	0	2	2
	Hypothesis 1:	0	1	1
	Hypothesis 1a (H1a):	0	1	1
	hypothesis was	0	1	1
	Hypothesis:	0	1	1
	hypothesize that	0	3	3
	hypothesized that	4	2	6
	registered ... hypotheses	1	0	1
test*		5	2	7
	test if	0	1	1
	test whether	1	1	2
	tested whether	2	0	2
	testing	1	0	1
	to ... test	1	0	1
test*, predict*	test ... prediction	0	1	1
predict*		4	0	4
	had ... predictions	1	0	1
	predicted that	2	0	2
	predicts that	1	0	1
examin*		5	0	5
	examine whether	2	0	2
	examined	1	0	1
	examined whether	1	0	1
	to examine	1	0	1
(other)		0	5	5
	(H1)	0	1	1
	expected that	0	1	1
	if ... then	0	1	1
	predication that	0	1	1
	we expect	0	1	1

Note. This table contains 44 hypothesis introduction phrases from 30 Registered Reports: 19 papers contributed one phrase each, 9 papers contributed two each, one contributed three, and one contributed four.

Table 3
Hypothesis introduction phrases in replication Registered Reports (testing previously studied hypotheses)

		source		
core word(s)	introduction phrase	abstract	full text	total
hypothes*, test*		2	1	3
	test ... hypotheses	0	1	1
	test ... hypothesis	1	0	1
	tested ... hypotheses	1	0	1
hypothes*, predict*	hypotheses predicted	1	0	1
hypothes*, examin*	examined ... hypothesis	1	0	1
hypothes*		2	5	7
	according to ... hypothesis	0	1	1
	Hypotheses	0	1	1
	Hypothesis 1 (H1):	0	1	1
	hypothesize that	0	1	1
	hypothesized that	2	1	3
test*		4	0	4
	testing whether	2	0	2
	to ... test	1	0	1
	to test	1	0	1
replicat*		20	3	23
	aim ... to replicate	0	1	1
	aim at replicating	1	0	1
	aimed to replicate	0	1	1
	attempted to replicate	1	0	1
	attempts to replicate	1	0	1
	conducted ... replication	3	0	3
	conducted ... replications	2	0	2
	performed ... replication	2	0	2
	present ... replication	1	0	1
	present ... replications	1	0	1
	replicated ... experiment	1	0	1
	replicating	0	1	1
	report ... replication attempt	1	0	1
	report ... replications	2	0	2
	sought to replicate	3	0	3
	we replicated	1	0	1
replicat*, examin*	critically examine and replicate	1	0	1
predict*	predicted that	2	0	2
examin*	examine whether	0	1	1
(other)		4	6	10
	establish whether	0	1	1
	H1	0	2	2
	investigate if	1	0	1
	sought to reproduce	1	0	1
	suggests that	2	0	2
	we ... conducted	0	1	1
	we assume	0	1	1
	we expect	0	1	1

Note. This table contains 53 hypothesis introduction phrases from 40 Registered Reports. One additional RR had no codeable hypothesis

5(4), e10068. doi:10.1371/journal.pone.0010068

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. doi:10.1016/j.ympev.2015.01.015

Rohatgi, A. (2018). *WebPlotDigitizer - Web Based Plot Digitizer*. Austin, Texas, USA. Retrieved from https://automeris.io/WebPlotDigitizer