

Social Psychology

Do Registered Reports Make Scientific Findings More Believable to the Public?

Elaine Costa¹ ^a, Yoel Inbar² , David Tannenbaum¹ 

¹ Department of Management, University of Utah, Salt Lake City, UT, USA, ² Department of Psychology, University of Toronto, ON, Canada

Keywords: scientific credibility, scientific rigor, transparency, open science, registered reports

<https://doi.org/10.1525/collabra.32607>

Collabra: Psychology

Vol. 8, Issue 1, 2022

Registered reports are an important initiative to improve the methodological rigor and transparency of scientific studies. One possible benefit of registered reports is that they may increase public acceptance of controversial research findings. We test this question by providing participants in a large US-based sample ($n = 1,500$) with descriptions of the key features of registered reports and the standard peer-review process, and then eliciting credibility judgments for various scientific results. We do not find evidence that participants view findings from registered reports as more credible than findings conducted under a standard (non-registered) report. This was true for both plausible and implausible study findings. Our results help clarify public attitudes and beliefs about scientific findings in light of recent methodological developments.

Scientific findings are often met with skepticism, especially when those findings defy widely-held beliefs. Notable examples include Barry Marshall demonstrating that the *H. pylori* bacterium causes ulcers, which challenged the prevailing view at the time that stress and lifestyle were the major causes of ulcer disease, and Francesco Redi's fly experiments offering evidence of biogenesis, which disputed the long-standing idea of spontaneous generation of living organisms (Azad, 2014; Gottdenker, 1979).

Science creates new knowledge by subjecting falsifiable hypotheses to empirical verification, but the accumulation of knowledge can only occur when individuals are willing to believe surprising results (i.e., if people are prepared to update their prior beliefs). While there are many reasons why people may be reluctant to believe a scientific finding, one potential concern is whether the researchers who produce a result are seen as honest and credible (Fiske & Dupree, 2014). Trust in scientists remains strong among the general public in the United States (Pew Research Center, 2019), but the recent replication crisis in the behavioral sciences threatens to erode that trust (Hendriks et al., 2020; Sarewitz, 2012; Vazire, 2017; Wagenmakers et al., 2012). The implication being that belief in a scientific result becomes more discretionary when a researcher or scientific field is viewed as less than fully credible. An open question is whether new research practices proposed in response to the replication crisis also help buffer or restore trust in scientific findings by the general public.

In this paper we focus on one such methodological reform: the adoption of registered reports (RRs). The essential feature of a registered report is that research proposals are

evaluated by reviewers *before* collection of data takes place. If the proposal is accepted, then journals commit to publishing the findings regardless of the results, so long as the research team faithfully carries out the agreed-upon study protocol (Hardwicke & Ioannidis, 2018). Rather than evaluating the worth of a paper once the results are known, publishing decisions under RRs are meant to focus on the importance of the research question, informativeness of the research design, and soundness of the proposed analyses. Proponents of RRs argue that this technique reduces the need to engage in questionable research practices (e.g., *p*-hacking) and eliminates publication bias, while also encouraging more informative and ambitious research designs (Kousta et al., 2020).

Given the assurance of publication irrespective of a study's results, registered reports are thought to be particularly well-suited to eliminate selective publication of study findings that favor successful outcomes (i.e., reporting positive rather than null results). Preliminary evidence suggests RRs are effective in this regard. A recent review in the field of psychology indicated that 96% of the traditional peer review publications sampled reported positive (i.e., hypothesis-confirming) results, compared to only 44% for RR publications (Scheel et al., 2021). Similarly, a review of 113 published registered reports in the biomedical and psychology fields indicated that approximately 60% of the hypotheses in the studies were not supported, compared to an estimated 5–20% of reported null findings in the traditional literature (Allen & Mehler, 2019). These studies indicate registered reports help mitigate publication bias in favor of positive findings.

^a Correspondence concerning this article should be addressed to Elaine Costa, email: elaine.costa@utah.edu

Beyond limiting publication bias, RRs may also improve study reproducibility and overall quality (Chambers & Tzavella, 2020). When a group of professional scientists were asked to rate the quality of various published studies without knowing which were registered reports, RR papers outperformed non-RR comparison papers on all 19 quality criteria (e.g., importance, rigor, novelty; Soderberg et al., 2020). Another investigation found mixed support that researchers were more likely to trust findings from RRs over unregistered studies (Field et al., 2020).

The modest evidence thus far indicates that the scientific community associates RRs with higher quality (i.e., more believable) work, but do such practices also influence beliefs of the general public? By increasing transparency, open science practices such as RRs reduce the information asymmetry between researchers and consumers of scientific output, and can thus alleviate uncertainty about the validity of a scientific finding (Vazire, 2017). For example, a recent Pew Research Center survey (2019) revealed that 57% of Americans report they trust scientific findings more when the data are openly available and 52% when the research has been independently reviewed. Second, by having a research plan vetted for soundness by peer reviewers prior to data collection, RRs can help to increase beliefs in the robustness of a finding. RRs may also communicate information about the researcher by serving as a genuine signal of a researcher's commitment to scientific transparency (Kraft-Todd & Rand, 2021). Thus, RRs may allow the general public to be more confident both in the quality and rigor of the research and the integrity of the researcher conducting the study.

We examine whether non-experts (i.e., members of the general public) are more willing to believe a scientific result published as a registered report, compared to when that same result comes from a conventional (non-registered report) scientific study. When describing RRs to participants, we restrict our descriptions of these methods to their fundamental features, and do not highlight the motivation or rationale for their implementation (e.g., to reduce publication bias or *p*-hacking). In other words, participants are only given information about the differences between RRs and non-RRs, such as their evaluation processes (i.e., assessment of the research methods take place before or after data is collected) and at which project stage journals commit to publishing the findings (i.e., as a proposal or a completed project). We also explore how RRs affect both willingness to believe unsurprising results (that do not require individuals to substantially update their prior beliefs) as well as surprising results (that do require greater revision of prior beliefs).

Study Overview

We investigate these questions in a pilot followed by a larger-scale study. In the pilot, we examine credibility ratings of registered reports relative to non-RRs, and whether any effects are moderated by the a priori plausibility of a scientific finding. We report the methods and results of the pilot in detail because our main study design closely parallels that used in the pilot study.

For both our pilot and main study, we determined our sample size in advance of data collection. For both studies we also pre-registered hypotheses and analysis plans. The completed pilot study and main study proposal were reviewed as a Stage 1 Registered Report, and received in-principle acceptance on August 11, 2021, prior to data collection for the main study. The Stage 1 manuscript (unchanged from the point of in-principle acceptance), as well as all study materials, data, and code can be found at <https://researchbox.org/154>.

Pilot Study

Method

We recruited 800 participants (50% male, mean age = 35.53 years, range: 18–84 years) to participate for \$0.80 each from the online labor market Prolific Academic. Before starting the study, participants were presented with a basic attention check (Oppenheimer et al., 2009). Those who failed the attention check were disqualified from continuing with the study.

Participants then read a short tutorial that explained the differences between the standard peer-review process and registered reports, which were described as “a novel way of publishing scientific findings.” Participants then responded to three multiple-choice comprehension questions, and those who failed to correctly answer all three questions were given a second chance and asked to try again. Participants who failed to correctly answer all three items a second time were disqualified from continuing on with the study. Exclusions due to inattention or failing comprehension questions took place before being randomly assigned to treatment conditions.

Each participant then read five scientific findings, randomly drawn from a pool of 10 vignettes. Vignettes were shown on separate pages in random order. For each vignette, participants were randomly assigned to one of four conditions that varied in two respects: (1) whether the outcome represented a plausible or implausible result based on pilot data,¹ and (2) whether the scientific finding was from a registered report. For instance, one vignette read:

1 We piloted ($N = 500$) twenty-five vignettes covering different topical areas to identify the vignettes that lay in an optimal “credibility area,” outside of those where strong prior beliefs from participants would leave no room for updating positions, but at the same time that the vignette was not so uncontroversial as to leave people indifferent to the outcome. In other words, those vignettes that allowed for sufficient flexibility in attitudes for the effects of the registered report manipulation to be discernible. For each vignette, we varied the outcome of the scientific finding, which we used to determine which scientific outcomes were perceived to be a priori plausible versus implausible. To do so, we recorded the absolute effect size in outcomes (using Cohen's d) for each vignette and then selected the 10 vignettes with a Cohen's d just below 1.00.

“With many countries legalizing marijuana, the possibility of cannabis leading to heavier drug use has been of public concern for some time. A recently published study is helping shed light on this issue. Researchers followed marijuana and non-marijuana users, and recorded various health-related habits over 15 years. Based on the results of this study, researchers concluded marijuana [is/is not] a “gateway drug.” Marijuana users [were/were not any] more likely than non-users to use other illegal drugs like cocaine or heroin.”

In this vignette, the finding that marijuana was revealed to be a gateway drug was rated by participants (in our pilot data) as more implausible than the finding that marijuana was not a gateway drug. Participants assigned to the RR condition also received the following text at the bottom of the vignette: “This study was published as a Registered Report. The journal committed to publishing this paper in advance before the results of the study were known.” Participants in the non-RR condition did not receive any additional text. An overview of the 10 vignettes is provided in Table 1 (note however that the descriptive statistics reported in Table 1 are from the main study).

After each vignette, participants were asked to rate the credibility of the finding by responding to four statements: “I find this study believable”, “I find this study convincing”, “I think this finding is likely to be false” and “I think it is likely this finding is not true.” All items were rated on 7-point scales (1 = *strongly disagree*, 7 = *strongly agree*). Responses to the last two items were reversed and we then averaged all responses to form a credibility index (Cronbach’s $\alpha = 0.96$).

As an exploratory measure, we measured beliefs about scientific bias at the end of the study. This item, taken from a Pew survey on the topic (Pew Research Center, 2019), asked participants to choose the statement closest to their opinion: “scientists make judgments based solely on the facts,” or “scientists’ judgments are just as likely to be biased as those of other people.” We coded participants as holding beliefs of scientific bias if they selected the latter statement.

Results

First, we examined whether RRs were viewed as more credible than non-RRs across all vignettes. Using OLS re-

gression,² we regressed credibility scores onto study outcome (0 = implausible outcome, 1 = plausible outcome) and registered report status (0 = no, 1 = yes). The model also included fixed effects for vignettes and participant-clustered standard errors. First, we found that plausible findings were rated as more credible than implausible findings, $b = 0.849$, $SE = 0.055$, $p < 0.001$. Second and more importantly, participants found RRs to be more credible than non-RRs, $b = 0.098$, $SE = 0.048$, $p = 0.040$. To provide an estimate of effect size, the coefficient for RRs was 11.5% of that found for study outcome (i.e., the expected difference in credibility ratings between plausible and implausible research findings), and represented a 0.07 standard deviation increase in overall credibility ratings based on the total variation observed in our sample. The point-biserial correlation between registered reports and credibility ratings was $r = 0.03$.

We also examined whether, at the vignette-level, the RR coefficient was systematically positive (i.e., that registered reports were generally viewed as more believable than non-registered reports). To do so, we coded whether the coefficient for registered reports was positive (0 = no, 1 = yes) for all 20 vignettes (10 scientific findings crossed by plausible vs implausible outcomes), and evaluated whether the proportion of positive coefficients was reliably larger than 50% using a binomial test.³ The coefficient for RR status was positive for 16 of the 20 vignettes ($p = 0.01$). Thus, the effect of learning that a study was a RR had a small but statistically significant impact on perceived credibility.

We next examined whether the effect of RRs on credibility was especially pronounced for a priori surprising study outcomes (i.e., do RRs help to close the “credibility gap” between plausible and implausible findings?). We used the same model specification as before, but now included an interaction term between study outcome and registered reports. The interaction effect between study outcome and registered report status was not statistically significant, $b = -0.044$, $SE = 0.088$, $p = 0.613$. Looking at average marginal effects in each condition, we find a positive but non-significant effect of RRs on credibility ratings for implausible study outcomes, $b = 0.120$, $SE = 0.071$, $p = 0.092$, and for plausible study outcomes, $b = 0.075$, $SE = 0.058$, $p = 0.192$. Thus, we fail to find statistically reliable evidence that RRs decrease the credibility gap between surprising and unsurprising results.

As an exploratory analysis, we also examined whether

2 For our pilot study we pre-registered a different analysis with the same outcome and predictor variables but which used a multi-level linear model with cross-random effects for participants and scenarios. Since conducting our pilot, we have been persuaded by recent recommendations to use OLS with clustered standard errors if one is simply trying to account for non-independence in the data (McNeish et al., 2017). We use this latter approach for the pre-registration of our main study and, for purposes of consistency, we report results using those same analyses for our pilot data. We note that using our original pre-registration plan returns similar results. When submitting credibility scores to a mixed effects model with study outcome and registered report as predictor variables, we find that plausible findings are viewed as more credible than implausible findings, $b = 0.864$, $SE = 0.096$, $p < 0.001$, and RRs as more credible than non-registered reports, $b = 0.099$, $SE = 0.044$, $p = 0.024$. When we apply the same model but also include an interaction term between study outcome and registered reports, we again find that the interaction effect is not statistically significant, $b = -0.056$, $SE = 0.087$, $p = 0.518$. We also pre-registered an additional interaction model that included both random slopes as well as random intercepts, but this model failed to converge (likely because the random variation around the slopes was close to zero). Finally, we again find no significant interaction effects between scientific bias beliefs and study outcomes or registered report status in the exploratory model, with p -values ranging between 0.373 and 0.826 for the two-way and three-way interactions. We report the full regression results from these models in Table S3 of the Supplementary Materials.

3 We did not pre-register this analysis.

beliefs about scientific bias moderate our results. We regressed credibility scores onto scientific bias beliefs (0 = no, 1 = yes), and again included scenario fixed effects and clustered standard errors by participants. Perhaps unsurprisingly, those who held beliefs of scientific bias also generally found scientific findings to be less credible, $b = -0.179$, $SE = 0.059$, $p = 0.003$. Next, to see whether science bias attitudes were sensitive to plausible vs. implausible study outcomes, we regressed credibility scores onto scientific bias beliefs, study outcome, and the interaction term between the two variables. We did not find a significant interaction effect, $b = 0.065$, $SE = 0.112$, $p = 0.562$. We then conducted a similar analysis but with registered reports, and again found a nonsignificant interaction term, $b = -0.064$, $SE = 0.101$, $p = 0.526$. Finally, we fit a model that included the three-way interaction between scientific bias beliefs, registered report status, and study outcomes. We again found no significant interaction effects between beliefs of scientific bias and study outcomes or registered report status (p -values ranged between 0.365 and 0.956 for all two-way and three-way interactions). Thus, we find that participants who viewed scientists as biased believed all scientific findings less, but do not find clear evidence that these beliefs moderated any of our results. Results for these regressions are reported in Table S2 of the Supplementary Materials.

Main Study

Our pilot data provides preliminary evidence that RR studies are viewed as more credible than non-RR studies by non-experts, although the effect size was modest and we did not find reliable evidence that RRs help to close the “credibility gap” between plausible and implausible research findings.

In the pilot study we described RRs as a “new” publishing method, which confounds publication method (RR vs. non-RR) with advancement from the status quo. It may be that participants would see any publication method that represents a change from the status quo as an improvement. To address this concern, in our main study we adopt a new manipulation that simply refers to RRs as “pre-study review” (to highlight that review decisions are made in advance of data collection) and refers to non-RRs as “post-study review” (to highlight that review decisions are made after data collection).

Method

Participants. We recruited 1,500 participants from Prolific Academic (46% male, mean age = 41.41 years, range: 18–92 years) to participate in a research study for \$0.70 each. Based on simulations⁴ that assume an effect size and within participant-clustering equal to that observed in our pilot data, this sample size provides us with 84% statistical power to detect an effect at $p \leq 0.05$.

Procedure. Our main study is identical to the methods used in our pilot study, but with the following key changes.

First, for the tutorial explaining the differences between the standard peer-review process and registered reports, and throughout the study, we referred to these publishing methods as “post-study review” and “pre-study review,” respectively. We also used neutral language so as to not indicate which method represented the status quo. Second, we provided information about non-RRs (rather than simply omitting information about RR status) for each vignette. Participants in the RR and non-RR conditions saw the following text at the bottom of each vignette: “This study was published using [pre-study review/post-study review]. The journal assessed the strengths and weaknesses of the study design [before/after] the results of the study were known.” Similar to our pilot study, we combined ratings into a credibility index for each vignette (Cronbach’s $\alpha = 0.95$). Table 1 reports descriptive statistics for each vignette.

Results

We first investigated whether RRs were viewed as especially credible using the same regression specification as in our pilot study. As shown in model 1 of Table 2, plausible findings were viewed as more credible than implausible findings, $b = 0.674$, $SE = 0.037$, $p < 0.001$. However, unlike our pilot results, this time RRs were not viewed as significantly more credible than non-RRs, $b = -0.025$, $SE = 0.042$, $p = 0.543$. In terms of effect size, the RR coefficient corresponded to approximately 4% of the effect found for study outcome, or a 0.02 standard deviation in total credibility scores across our sample, with the point-biserial correlation between registered reports and credibility ratings being 0.01. Thus, we did not find support for the hypothesis that participants view RRs as more credible than non-RRs.

We again examined whether, at the vignette-level, the RR version was generally viewed as more credible than non-registered reports. We coded the number of positive RR coefficients (0 = no, 1 = yes) for the 20 vignette-outcome combinations and tested against a null of 50% using a binomial test. Exactly half of the 20 vignettes had positive coefficients, which unsurprisingly is consistent with the null hypothesis of no difference in credibility ratings as a function of RRs ($p > 0.99$). As depicted in Figure 1, the range in effect sizes for RRs (Cohen’s d) across vignettes ranged from -0.17 to 0.27 . Thus, when analyzing data at the vignette-level, rather than at the trial-level, we again fail to find support for the hypothesis that RRs are viewed as more credible than non-RRs.

We next examined whether the impact of RRs on credibility ratings was particularly effective for unexpected scientific results (Table 2, model 2). Consistent with the results from our pilot study, the interaction between RRs and study outcome was not significant, $b = -0.042$, $SE = 0.062$, $p = 0.502$. Thus, we do not find that RRs, compared to non-RRs, help reduce differences in credibility scores between plausible and implausible scientific findings.

We also examined whether our effects were moderated by beliefs about scientific bias. We first regressed credibility

⁴ Code for our power analysis simulations can be found at: <https://researchbox.org/154>.

Table 1. Means and Standard Deviation by Condition

Scientific Finding	Registered Report				Non-Registered Report			
	Plausible		Implausible		Plausible		Implausible	
	M	SD	M	SD	M	SD	M	SD
Atheists and agnostics were [less/no more biased] against gender fluid individuals than those with a religious belief system.	5.22	1.27	4.34	1.51	5.17	1.15	4.57	1.40
Couples with children are [just as likely/more likely] to cheat.	4.79	1.39	4.46	1.33	4.89	1.22	4.48	1.34
Compared to non-prisoners, ex-prisoners are [no less/less] reliable and trustworthy at work.	5.15	1.15	4.37	1.34	5.00	1.30	4.52	1.40
Fake news [helped/did not help] elect Donald Trump.	4.58	1.69	4.34	1.50	4.81	1.54	4.08	1.57
Gender discrimination was [found/not found] in small business lending.	5.07	1.35	4.32	1.50	5.04	1.32	4.24	1.50
Marijuana users were [not any more likely/more likely] than non-users to use other illegal drugs like cocaine or heroin.	5.27	1.26	4.35	1.67	5.18	1.22	4.55	1.59
Raising the minimum wage [does not lead/leads] to increased unemployment.	5.02	1.34	4.31	1.65	5.17	1.28	4.20	1.41
Greater social media use [leads/does not lead] to depression and reduced well-being.	5.04	1.24	4.12	1.51	5.35	1.08	4.15	1.44
Math and science scores were found on average to be [the same/lower] for female students.	5.20	1.32	4.33	1.43	5.20	1.34	4.21	1.46
Playing violent video games [has no effect on/increases] aggressive behavior in adolescents.	4.87	1.42	4.71	1.57	4.83	1.42	4.70	1.33

Note: Description of scientific findings are abridged versions of the ones viewed by participants, which also included a brief topic introduction and methodological description of the study. Credibility scores of study findings were rated on 7-point scales (1 = strongly disagree, 7 = strongly agree). The first outcome listed in brackets for each finding represents the plausible condition, while the second represents the implausible condition.

scores onto scientific bias (0 = no scientific bias, 1 = scientific bias), and again found that participants who viewed scientists as biased were more inclined to regard scientific findings as generally less credible than those who do not view scientists as biased, $b = -0.206$, $SE = 0.042$, $p < 0.001$ (Table 2, model 3). We next regressed credibility scores onto scientific bias beliefs, study outcome, and the interaction between the two variables (Table 2, model 4). The interaction effect was positive but not statistically significant, $b = 0.127$, $SE = 0.074$, $p = 0.087$. Looking at the average marginal effects, the effect of study outcome (plausible vs implausible findings) was more pronounced among participants who view scientists as biased, $b = 0.743$, $SE = 0.056$, $p < 0.001$, than among participants who do not view scientists as biased, $b = 0.616$, $SE = 0.049$, $p < 0.001$. In model 5 of Table 2, we regressed credibility scores onto scientific bias beliefs, registered reports, and the interaction between the two variables. We found a positive and significant interaction effect between scientific bias and RRs, $b = 0.395$, $SE = 0.085$, $p < 0.001$. Looking at the average marginal effects, the effect of registered reports (RRs vs non-RRs) increased the credibility of a study finding for participants who view

scientists as biased, $b = 0.195$, $SE = 0.064$, $p = 0.002$, but reduced the credibility of scientific findings for participants who do not view scientists as biased, $b = -0.200$, $SE = 0.055$, $p < 0.001$. Lastly, as shown in model 6 of Table 2, we do not find a reliable three-way interaction between scientific bias beliefs, study outcome, and RRs, $b = -0.101$, $SE = 0.124$, $p = 0.416$.

In the Supplemental Materials we also report an analysis that aggregates results from the pilot and main study. Although such aggregate analyses should be interpreted with caution given the differences in design between our pilot and main study, we find largely similar results to those reported from our primary analysis.

Discussion

Registered reports represent a promising new initiative for improving rigor and transparency in empirical research. Our research question, and the title of this article, asked whether registered reports make scientific findings more believable to the general public. Confirming Hinchliffe's rule,⁵ we do not find support for the hypothesis that study findings from RRs are viewed as generally more credible

Table 2. Effect of Registered Reports and Scientific Bias on Scientific Findings' Credibility

	(1)	(2)	(3)	(4)	(5)	(6)
Study Outcome	0.674*** (0.037)	0.695*** (0.048)		0.616*** (0.049)		0.607*** (0.063)
Registered Report	-0.025 (0.042)	-0.005 (0.055)			-0.200*** (0.055)	-0.195*** (0.074)
Scientific Bias			-0.206*** (0.042)	-0.260*** (0.057)	-0.406*** (0.058)	-0.477*** (0.077)
Registered Report x Study Outcome		-0.042 (0.062)				0.012 (0.084)
Scientific Bias x Study Outcome				0.127 (0.074)		0.180 (0.095)
Scientific Bias x Registered Reports					0.395*** (0.085)	0.426*** (0.110)
Scientific Bias x Study Outcome x Registered Reports						-0.101 (0.124)
Intercept	4.501*** (0.056)	4.490*** (0.059)	4.918*** (0.055)	4.604*** (0.058)	5.016*** (0.059)	4.702*** (0.066)
Vignette Fixed Effects	yes	yes	yes	yes	yes	yes
Observations	7,500	7,500	7,500	7,500	7,500	7,500
Participants	1,500	1,500	1,500	1,500	1,500	1,500
Vignettes	10	10	10	10	10	10
R ²	0.059	0.059	0.010	0.064	0.015	0.069

Note: Columns correspond to OLS regression coefficients, with participant-clustered standard errors in parentheses. The dependent variable in all models is an index of credibility judgments scored on a 7-point scale, with positive values denoting higher credibility judgments. Study Outcome takes on the value of 0 if the scientific finding was rated as implausible and 1 if the scientific finding was rated as plausible, based on pilot data. Registered Report takes on the value of 1 for the presence of a registered report, and 0 for a non-registered report. Scientific Bias takes on the value of 1 for the presence of scientific bias beliefs and 0 for its absence. For scenarios, we dummy-coded 10 vignettes with the "Atheists/Agnostics" scenario representing the reference value. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

than non-RRs. Earlier studies suggest that registered reports are generally associated with greater rigor and quality by individuals in the scientific community (Chambers & Tzavella, 2020; Soderberg et al., 2020). The present study extends this question to the public at large, and finds that these perceptions are not currently reflected among the general population. RRs also do not appear to help to close the "credibility gap" between plausible and implausible findings.

In an exploratory analysis, we do find that RRs, compared to non-RRs, increase credibility in scientific results among those who report skepticism of scientists, while having inconsistent effects on those who believe scientists make judgments based solely on the facts. If future research is able to replicate this finding, it suggests that RRs may enhance the credibility of scientific outcomes among the subset of participants most inclined to dismiss such results. Practices that can enhance trust in scientific findings among those most skeptical of the scientific establishment may become even more critical given the continued polarization and politicization surrounding scientific findings (Lee, 2021; Parikh, 2021; Rekker, 2021).

Naturally, our study has several limitations. Our sample consisted of participants from an online labor market who

are unrepresentative of the broader population in some respects. For instance, participants on online research platforms tend to be more educated than the U.S. population (e.g., Peer et al., 2017). Sample non-representativeness should be kept in mind when drawing inferences about how the general public would react to scientific information. Another limitation is the somewhat stylized nature of our study design. Participants in our studies first completed a tutorial explaining the difference between RRs and standard peer review studies before rating scientific findings for their credibility. To isolate the causal effect of RR formats on credibility judgments, we limited the descriptions of these publishing methods to their fundamental features (e.g., decisions to publish are made before or after data has been collected and reported), without alluding to their benefits or providing a rationale for why registered reports have been developed as an alternative to the traditional publication process (e.g., to prevent p -hacking and publication bias). Although our tutorial was necessary to ensure that participants properly understood the construct, it is unlikely that most public consumers of scientific information will be as familiar with RRs. On one hand, the salience of the tutorial and manipulation in this study could mean our results represent an upper bound on how RRs may increase

⁵ Hinchcliffe's rule (attributed to the physicist Ian Hinchcliffe) states that if the title of a scholarly article is a yes-no question, the answer to that question is "no" (Shieber, 2015).

Do Registered Reports Make Scientific Findings More Believable to the Public?

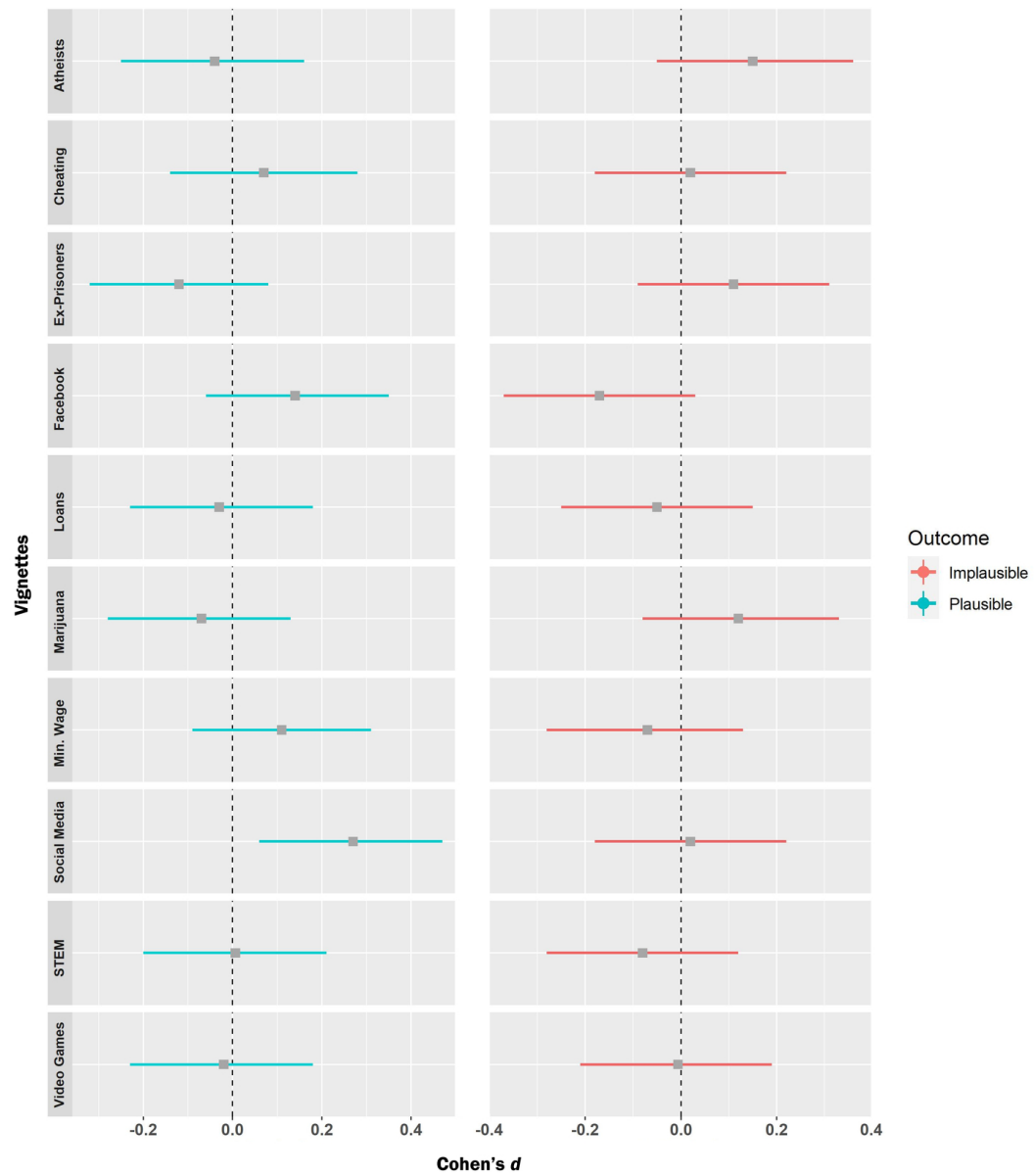


Figure 1. Registered Report Effect Sizes by Vignette and Study Outcome (95% Confidence Interval)

Effects of registered reports versus non-registered reports on credibility ratings with corresponding 95% confidence intervals per vignette and study outcome.

the credibility of scientific findings. On the other hand, our findings may represent a conservative estimate of the influence of RRs on the believability of scientific results, given that RRs are still relatively unfamiliar to the public. An interesting avenue for future research is to examine whether the credibility of scientific results changes over time as the general public becomes more broadly familiar with registered reports.

When individuals are exposed to new information, there are generally two major dimensions of credibility judgments: the extent to which an audience believes the message and the extent to which an audience believes the messenger (Roberts, 2010). In this study, we examined how RRs affect the credibility of the message being relayed. Future research may wish to explore how the information about the adoption of registered reports influences the perceived credibility of the scientific community at large (i.e., the messenger) rather than the message itself.

Besides questions concerning how to best convey information about new scientific practices, future research could examine how RRs compare to other existing open practice initiatives (such as publicly-available data or pre-registrations “badges”; Kidwell et al., 2016) in improving the credibility of a finding. While existing practices such as badges largely improve transparency, RRs signal both transparency and rigor. Future studies may wish to explore whether the general public differentiates between these two factors, so

that scientific practices can be better aligned with the high standards the public expects from scientific research.

Author Contributions

Contributed to conception and design: EC, DT, and YI
 Contributed to acquisition of data: EC, DT, and YI
 Contributed to analysis and interpretation of data: EC, DT, and YI
 Drafted and/or revised the article: EC, DT, and YI
 Approved the submitted version for publication: EC, DT, and YI

Competing Interests

The authors declare they have no competing interests. YI is a senior editor at Collabra: Psychology. He was not involved in the review process of this manuscript.

Data Accessibility Statement

The full dataset, materials and syntax used in this paper are available on ResearchBox at:
<https://researchbox.org/154>

Submitted: June 04, 2021 PST, Accepted: January 31, 2022 PST



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

- Allen, C., & Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PLoS Biology*, 17(5), e3000246. <https://doi.org/10.1371/journal.pbio.3000246>
- Azad, M. (2014). Q&A: Barry Marshall. *Nature*, 514(7522), S6–S7. <https://doi.org/10.1038/514s6a>
- Chambers, C. D., & Tzavella, L. (2020). The past, present, and future of Registered Reports. *MetaArXiv*. <https://doi.org/10.31222/osf.io/43298>
- Field, S. M., Wagenmakers, E.-J., Kiers, H. A. L., Hoekstra, R., Ernst, A. F., & van Ravenzwaaij, D. (2020). The effect of preregistration on trust in empirical research findings: Results of a registered report. *Royal Society Open Science*, 7(4), 181351. <https://doi.org/10.1098/rsos.181351>
- Fiske, S. T., & Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics. *Proceedings of the National Academy of Sciences*, 111(Supplement 4), 13593–13597. <https://doi.org/10.1073/pnas.1317505111>
- Gottdenker, P. (1979). Francesco Redi and the fly experiments. *Bulletin of the History of Medicine*, 53(4), 575–592.
- Hardwicke, T. E., & Ioannidis, J. P. A. (2018). Mapping the universe of registered reports. *Nature Human Behaviour*, 2(11), 793–796. <https://doi.org/10.1038/s41562-018-0444-y>
- Hendriks, F., Kienhues, D., & Bromme, R. (2020). Replication crisis = trust crisis? The effect of successful vs failed replications on laypeople's trust in researchers and research. *Public Understanding of Science*, 29(3), 270–288. <https://doi.org/10.1177/0963662520902383>
- Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology*, 14(5), e1002456. <https://doi.org/10.1371/journal.pbio.1002456>
- Kousta, S., Pastrana, E., & Swaminathan, S. (2020). Three approaches to support reproducible research. *Science Editor*, 42(3), 77–82.
- Kraft-Todd, G. T., & Rand, D. G. (2021). Practice what you preach: Credibility-enhancing displays and the growth of open science. *Organizational Behavior and Human Decision Processes*, 164, 1–10. <https://doi.org/10.1016/j.obhdp.2020.10.009>
- Lee, J. J. (2021). Party Polarization and Trust in Science: What about Democrats? *Socius*. <https://doi.org/10.1177/23780231211010101>
- McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, 22(1), 114–140. <https://doi.org/10.1037/met0000078>
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4), 867–872. <https://doi.org/10.1016/j.jesp.2009.03.009>
- Parikh, S. (2021). *Why we must rebuild trust in science*. PEW Charitable Trust.
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153–163. <https://doi.org/10.1016/j.jesp.2017.01.006>
- Pew Research Center. (2019). *Trust and Mistrust in Americans' Views of Scientific Experts*. https://www.pewresearch.org/science/wp-content/uploads/sites/16/2019/08/PS_08.02.19_trust.in_scientists_FULLREPORT-1.pdf
- Rekker, R. (2021). The nature and origins of political polarization over science. *Public Understanding of Science*, 30(4), 352–368. <https://doi.org/10.1177/0963662521989193>
- Roberts, C. (2010). Correlations among variables in message and messenger credibility scales. *American Behavioral Scientist*, 54(1), 43–56. <https://doi.org/10.1177/0002764210376310>
- Sarewitz, D. (2012). Beware the creeping cracks of bias. *Nature*, 485(7397), 149.
- Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An excess of positive results: Comparing the standard Psychology literature with Registered Reports. *Advances in Methods and Practices in Psychological Science*, 4(2), 1–12. <https://doi.org/10.1177/25152459211007467>
- Shieber. (2015). Is this article consistent with Hinchliffe's Rule? *Annals of Improbable Research*, 21(3), 18–19.
- Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J. G., Singleton Thorn, F., Vazire, S., Esterling, K., & Nosek, B. A. (2020). Initial Evidence of Research Quality of Registered Reports Compared to the Traditional Publishing Model. *MetaArXiv*. <https://doi.org/10.31222/osf.io/7x9vy>
- Vazire, S. (2017). Quality Uncertainty Erodes Trust in Science. *Collabra: Psychology*, 3(1), 1–5. <https://doi.org/10.1525/collabra.74>
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science*, 7(6), 632–638. <https://doi.org/10.1177/1745691612463078>

SUPPLEMENTARY MATERIALS

Peer Review History

Download: https://collabra.scholasticahq.com/article/32607-do-registered-reports-make-scientific-findings-more-believable-to-the-public/attachment/82366.docx?auth_token=FtjOdYpkCB0qkCpNJQK-

Supplemental Material

Download: https://collabra.scholasticahq.com/article/32607-do-registered-reports-make-scientific-findings-more-believable-to-the-public/attachment/82367.docx?auth_token=FtjOdYpkCB0qkCpNJQK-
