METASCIENCE

Research on registered report research

Can a publication format shape qualities of published research? Higgs and Gelman discuss a new study comparing peer-reviewers' perceptions of Registered Reports to those of standard research articles. The authors conclude the registered publications were at least as good on the qualities measured, and they discuss challenges of doing research on research.

Megan D. Higgs and Andrew Gelman

Good research is hard. Good research on research is probably harder. Not only does it carry all the challenges of lack of control and difficulties in measurement found in social science research, but by its subject matter invites a greater degree of scrutiny and criticism. A new article by Soderberg et al.[1] in *Nature Human Behaviour* compares perceptions of research qualities of registered reports to papers published using the traditional model. Registered reports require peer review of the research, and acceptance by the journal, before data are collected and analyzed—and thus before results are reported—thereby shifting the focus of review away from conclusions and onto the proposed methods and underlying theory. This should counter the problem of selecting studies for publication based on statistical results (such as *P* values), while potentially ignoring methodological or theoretical shortcomings. Critics of widespread adoption of registered reports worry that the practice could favor a more a mechanistic process of science and stifle creativity, novelty, and discovery.

In their study, Soderberg et al. selected pairs of publications—one coming through the registered report model and the other coming through the standard publication model—and had volunteer researchers evaluate them on 19 properties on a −4 to +4 scale, where 0 represents the score for an average paper. They summarize the results as mean differences in scores within pairs of papers, and as their article title ("Initial evidence of research quality of registered reports compared to the standard publishing model") suggests, they focus most on the question of 'research quality'. It seems likely that consumers of such research implicitly assume a particular level of research quality for a published study on research quality—and this puts an added responsibility on meta-research researchers, as well as reviewers and readers.

Soderberg et al.[1] have measurement, analysis, and interpretation details that we, as statisticians, could spend ink quibbling about. The lack of such quibbling shouldn't be taken as implicit agreement with those details. Here, though, we're focused on the bigger picture. Have we have learned something meaningful about qualities of registered reports relative to standard publications? Is this study helpful to our collective understanding of research practice and arguments for reform? Soderberg et al.[1] will likely be cited often, as support for various arguments regarding the effectiveness of registered reports in research, and most notably for the increase they found in research quality.

It's interesting, and probably important, to consider prior beliefs about the benefits and detriments of a registered reports model. What would the takeaway be if this study had found that mean scores for research quality were lower for registered reports as compared to the standard model? Considering the study from that (hypothetical) unexpected outcome raises some obvious questions, such as 'how was research quality measured?' and 'how large was the observed difference and is it practically relevant?' We suspect a lot of time would be spent addressing these questions for an unexpected outcome, but less so for the expected outcome—making the thought exercise provide an interesting reflection for scientists in general, a sneaky manifestation of confirmation bias.

When does knowing (or assuming) a paper was a registered report become an indicator of perceived research quality itself? And how does this complicate assessment and interpretation of results? Soderberg et al.[1] acknowledge and look into this. Indeed, their data seem to support the existence of some confirmation bias, with slightly greater differences in scores (on average) reported from reviewers who reported believing the registered report format improves rigor and quality. It's hard to tease everything apart, and this is a great example of nuances that researchers studying new research practices need to consider.

With all this in mind, it seems more interesting to look at qualities of research previously suggested to deteriorate under the registered reports model. Would widespread adoption of registered reports contribute to boring science and possibly reduce creativity of research in general? Soderberg et al.[1] included measures of creativity and innovation in their study and concluded that "the standard model did not outperform [registered reports] for any outcome" based on average scores over all pairs of papers and all reviewers (and of course conditional on the papers and reviewers included in the study, the instrument used for measurement, analysis, and so on). This conclusion tends towards the dangerous territory of taking lack of evidence as no evidence, but does provide more than anecdotal information about the potential worrisome implications of registered reports on research in general. All this leads us to take the study results as support for the claim that registered reports do not make research worse, but the study does not take us all the way to evidence that registered reports improve research quality.

Underlying any of the conclusions are the substantial challenges of measurement: defining what qualities to include in the survey, what scale to use for each, and how to interpret changes in the scale relative to what is meaningful in practice. These challenges are inherent in social science research, and research on research is stuck with them.

Choosing which qualities to measure in the context gets at the deep question of what makes research high-quality (or highly credible if that's what we are after) across studies and disciplines. There seems to be an implicit assumption in science reform that there is more agreement on this than perhaps there really is in practice. Research on research depends on these definitions and could play a unique role in pushing these ideas forward or at least in modelling

meaningful discussion and critique about the theory underlying the research ideas and particularly in how it relates to decisions about how to measure qualities in research. Does the scientific community agree on the most important characteristics of research quality or research credibility, as well as on how to measure them? Individual researchers have different standards and tend to focus on different aspects of quality, not to mention the differences that exist among disciplines. This is just something we should have in mind.

Even if we all agree on the set of qualities to evaluate, there is still a big leap from that to creating a measurement instrument and working through difficulties in interpretation relative to its scale. For example, what should we take away from Soderberg and colleagues' reported estimate of 0.39 (95% credible interval from 0.03 to 0.76) for the difference in perceived research quality between registered reports and standard publications? We would love to see arguments for and against taking a magnitude such as 0.39 as practically meaningful to research practice or science reform!

One reason we think that Soderberg et al.[1] holds up is that the authors are ultimately only trying to argue that registered reports are no worse than current standard practice. It's hard to make a strong positive claim from this sort of observational study, given the potential for selection bias in what sorts of studies are deemed worth the effort of a registered report in the first place. But we are convinced by the weaker claim, that given these data it would be hard to make the case that registered reports make things worse. 'Not worse' is not as exciting as 'better', but it still represents a step forward in assessing the potential consequences of registered reports and other proposed science reforms. ❐

Megan D. Higgs [1]✉ and Andrew Gelman [2]✉

[1]*Critical Inference LLC, Bozeman, MT, USA.* [2]*Department of Statistics and Department of Political Science, Columbia University, New York, NY, USA.* ✉*e-mail: megan.higgs@critical-inference.com; gelman@stat.columbia.edu*

Published online: 24 June 2021
https://doi.org/10.1038/s41562-021-01148-y

References

1. Soderberg, C. K., et al. *Nat. Hum. Behav.* https://doi.org/10.1038/s41562-021-01142-4 (2021).

Competing interests

The authors declare no competing interests.