Preregistration is redundant, at best

Aba Szollosi[1]*, David Kellen[2], Danielle J. Navarro[1], Richard Shiffrin[3], Iris van Rooij[4], Trisha Van Zandt[5], and Chris Donkin[1]

[1] University of New South Wales

[2] Syracuse University

[3] Indiana University

[4] Radboud University

[5] Ohio State University

Keywords: preregistration, theory-development, inference

Acknowledgements

*Correspondence: aba.szollosi@gmail.com

The key implication argued by proponents of preregistration is that it improves the diagnosticity of statistical tests [1]. In the strong version of this argument, preregistration does this by solving statistical problems, such as family-wise error rates. In the weak version, it nudges people to think more deeply about their theories, methods, and analyses. We argue against both: the diagnosticity of statistical tests depend entirely on how well statistical models map onto underlying theories, and so improving statistical techniques does little to improve theories when the mapping is weak. There is also little reason to expect that preregistration will spontaneously help researchers to develop better theories (and, hence, better methods and analyses).

Fixing science with statistics

Scientific progress consists of iterative attempts to improve theories by careful attempts to find and correct errors in them [2,3]. One useful technique for this process is to describe theories mathematically. In the social and behavioural sciences, such mathematical formalisms are typically realized by statistical models. Yet the extent to which such models are useful depends entirely on how accurate the match between the theory and model is (see Box 1). Because the match between theories and statistical models is generally poor in psychological science, scientific conclusions based on statistical inference are often weak [4,5].

Solving statistical problems with preregistration does not compensate for weak theory. Imagine making a random prediction regarding the outcome of an experiment. Should we observe the predicted outcome, we would not regard this “theory” as useful for making subsequent predictions. Why should we regard it as better if it was preregistered? Researchers can preregister bad theories, the predictions of which are no better than randomly picking an outcome, but which can nonetheless bear out in experiments.

On the other hand, unlike statistical analyses, there is nothing inherently problematic about post-hoc scientific inference when theories are strong. The crucial difference is that strong scientific inference requires that post-hoc explanations are tested just as rigorously as ones generated before an experiment -- for example, by a collection of post-hoc tests that evaluate the many regularities implied by a novel theory [6]. There is no reason not to take such post-hoc theories seriously just because they were thought of after or were not preregistered before an experiment was conducted.

Preregistration as a nudge?

Although preregistration does not require the improvement of theories, many argue that it at least nudges researchers to think more deeply about how to improve their theories. Though this might sometimes be so, there is no clear explanation for why we should expect it to happen.

One possible explanation is that researchers are motivated to improve their theories should they encounter problems when preregistering a study or when preregistered predictions are not observed. The problem with this line of argument is that any improvement depends upon a good understanding of how to improve a theory, and preregistration provides no such understanding. In other words, although preregistration may prompt more thinking, it is not clear how it provides a path towards better thinking. The danger is that, without this understanding, preregistration may be perceived as the solution.

Instead, better understanding of theory development requires thoughtful discussion and debate about what constitutes good scientific inference. For example, what are the best examples of good theories both within and outside of the social and behavioural sciences [7]? What are characteristics of good theories [6]? How do we improve the link between psychological theory, measurement, methodology, and statistics? The answers to these questions are unlikely to come from nudging researchers with preregistration or some other method-oriented solution. They are likely to come from scientific problem-solving: generating, exchanging and criticizing possible answers, and improving them when needed.

Preregistration is redundant, and potentially harmful

Preregistration can be hard, but it can also be easy. The hard work associated with good theorizing is independent of the act of preregistration. What matters is the strength of the theory upon which scientific inference rests. Preregistration does not require that the underlying theory be strong, nor does it discriminate between experiments based on strong or weak theory.

Taking preregistration as a measure of scientific excellence can be harmful, because bad theories, methods, and analyses can also be preregistered. Requiring or rewarding the act of preregistration is not worthwhile when its presumed benefits can be achieved without it just as well. As a field, we have to grapple with the difficult challenge of improving our ability to test theories, and so should be wary of any 'catch-all' solutions like preregistration.

Box 1: Scientific vs statistical inference

Because statistical inference is so often used to inform scientific inference, it is easy to conflate the two. However, they have fundamentally different aims.

Scientific inference is the process of developing better theories. One aspect of theory development is to test whether the implications of a theory are observed in experiments. We can use statistics to help decide whether such implications are realized. Used in this way, statistical inference is the reliance on (strong, and false) mathematical assumptions to simplify a problem to permit scientific inference.

Some problems arising from statistical assumptions, such as family-wise error rates, only exist when hypotheses and statistical comparisons are effectively chosen at random. When statistical inference is used in scientific argument, statistical methods are just tools to test implications derived from theory. Therefore, such statistical problems become irrelevant because theories, not random selection, dictate what comparisons are necessary [8].

Problems in scientific inference, on the other hand, are important when the overall goal is theory development. Issues that prevent criticism of theory, such as poor operationalization, imprecise measurement, and weak connection between theory and statistical methods, need our attention instead of problems with statistical inference.

References

1 Nosek, B.A. *et al.* (2019) Preregistration Is Hard, And Worthwhile. *Trends Cogn. Sci.* 23, 815–818

2 Popper, K. (1959) *The Logic of Scientific Discovery*, Routledge.

3 Deutsch, D. (2011) *The Beginning of Infinity: Explanations that Transform the World*, Allen Lane.

4 Navarro, D.J. (2018) Between the Devil and the Deep Blue Sea: Tensions Between Scientific Judgement and Statistical Model Selection. *Comput. Brain Behav.* 2, 28–34

5 Szollosi, A. and Donkin, C. (2019) Neglected Sources of Flexibility in Psychological Theories: from Replicability to Good Explanations. *Comput. Brain Behav.* 2, 190–192

6 Szollosi, A. and Donkin, C. (2019) Arrested theory development: The misguided distinction between exploratory and confirmatory research. DOI: 10.31234/osf.io/suzej

7 Fiedler, K. (2018) The Creative Cycle and the Growth of Psychological Science. *Perspect. Psychol. Sci.* 13, 433–438

8 Oberauer, K. and Lewandowsky, S. (2019) Addressing the theory crisis in psychology. *Psychon. Bull. Rev.* 26, 1596–1618