

Stephen Senn

Edinburgh

The replication crisis: are P-values the problem and are Bayes factors the solution?

Today's posterior is tomorrow's prior. Dennis Lindley¹ (P2)

It has been claimed that science is undergoing a replication crisis and that when looking for culprits, the cult of significance is the chief suspect. It has also been claimed that Bayes factors might provide a solution.

In my opinion, these claims are misleading and part of the problem is our understanding of the purpose and nature of replication, which has only recently been subject to formal analysis². What we are or should be interested in is truth. Replication is a coherence not a correspondence requirement³ and one that has a strong dependence on the size of the replication study⁴.

Consideration of Bayes factors raises a puzzling question. Should the Bayes factor for a replication study be calculated as if it were the initial study? If the answer is yes, the approach is not fully Bayesian and furthermore the Bayes factors will be subject to exactly the same replication 'paradox' as P-values. If the answer is *no*, then in what sense can an initially found Bayes factor be replicated and what are the implications for how we should view replication of P-values?

A further issue is that little attention has been paid to false negatives and, by extension to true negative values. Yet, as is well known from the theory of diagnostic tests, it is meaningless to consider the performance of a test in terms of false positives alone.

I shall argue that we are in danger of confusing evidence with the conclusions we draw and that any reforms of scientific practice should concentrate on producing evidence that is reliable as it can be *qua* evidence. There are many basic scientific practices in need of reform. Pseudoreplication⁵, for example, and the routine destruction of information through dichotomisation⁶ are far more serious problems than many matters of inferential framing that seem to have excited statisticians.

References

1. Lindley DV. *Bayesian statistics: A review*. SIAM; 1972.
2. Devezer B, Navarro DJ, Vandekerckhove J, Ozge Buzbas E. The case for formal methodology in scientific reform. *R Soc Open Sci*. Mar 31 2021;8(3):200805. doi:10.1098/rsos.200805
3. Walker RCS. Theories of Truth. In: Hale B, Wright C, Miller A, eds. *A Companion to the Philosophy of Language*. John Wiley & Sons,; 2017:532-553:chap 21.
4. Senn SJ. A comment on replication, p-values and evidence by S.N.Goodman, *Statistics in Medicine* 1992; 11:875-879. Letter. *Statistics in Medicine*. 2002;21(16):2437-44.
5. Hurlbert SH. Pseudoreplication and the design of ecological field experiments. *Ecological monographs*. 1984;54(2):187-211.
6. Senn SJ. Being Efficient About Efficacy Estimation. Research. *Statistics in Biopharmaceutical Research*. 2013;5(3):204-210. doi:10.1080/19466315.2012.754726