

# A universe of uncertainty hiding in plain sight

Letter forthcoming in the *Proceedings of the National Academy of Sciences*

Per Engzell\*

December 2022

## Abstract

A recent study by Breznau et al. [“Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty” *PNAS*, 119(44) (2022)] raises concerns about the reliability of social science. I raise four concerns about its interpretation: 1) The study tests not one but several hypotheses; 2) The study successfully replicates a null finding; 3) Much variation is within results from a given team; 4) The data are inadequate for the hypothesis. Progress in social science requires attention to theory, measurement and causal inference in addition to variability of results.

In an impressive effort, Breznau et al. (1) (henceforth, BRW) report a many-analysts collaboration where multiple teams were involved in analyzing the same data and hypothesis: that immigration undermines public support for social policy. Like other such studies (2,3), the results show considerable variation and are sure to ignite debate. The message is clear: social scientists must be principled about analytical choices, transparent about their data and procedures, and humble about uncertainty. Analyses such as this have much to teach us about advancing those aims.

Many readers will find BRW’s “hidden universe of uncertainty” harrowing and ask if we should trust social science at all. Such an implication would be overwrought, for several reasons.

The study tests not one but several hypotheses. BRW collected 2 measures of immigration (4 including alternative sources) and 6 measures of policy support. The outcome variables span various domains: jobs, healthcare, pensions, unemployment, redistribution, and housing. Tested models include various combinations of within- and between-country variation, different countries and years, and so on. Given the variety of implied hypotheses, it would be remarkable if results did not vary.

Despite this variation, the study successfully replicates a published null finding. BRW chose their research question because it is “influential, long standing, and typical” (1). But according to the earlier study that BRW replicate (4), results “mostly fail to support” the hypothesis. BRW do not correct for multiple hypothesis testing, but most of their results span a narrow range around zero. As a replication, it is not clear that this should count as a failure and it may well count as a success.

---

\*UCL Social Research Institute, University College London; Nuffield College, University of Oxford; and Swedish Institute for Social Research, Stockholm University. E-mail: [p.engzell@ucl.ac.uk](mailto:p.engzell@ucl.ac.uk)

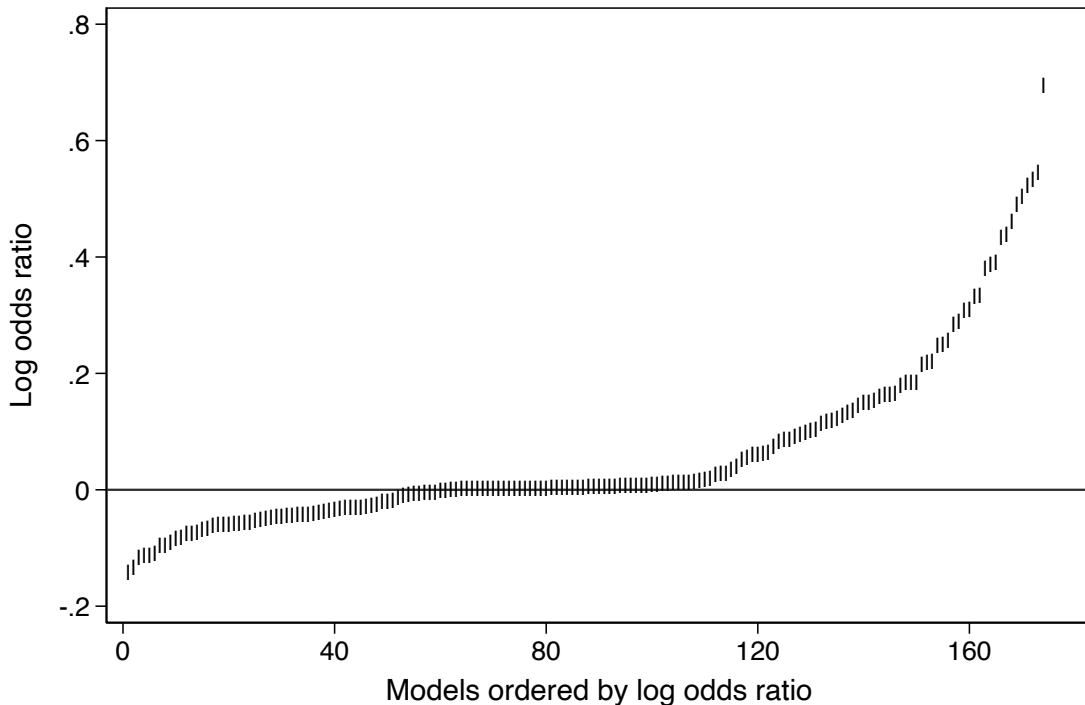


Figure 1: Variation in reported results from Brady and Finnigan (4) which Breznau et al. (1) replicate. Point estimates for 3 variables: percent foreign born, net migration, and change in percent foreign born. Total number of estimates  $n=174$ . Just like in Breznau et al., results span coefficients that are negative, positive, and indistinguishable from zero. Positive log odds ratios imply an association opposite of the hypothesized negative effect of immigration on support for social policy. Odds ratios as reported by Brady and Finnigan are not directly comparable with average marginal effects as reported by Breznau et al.

Much variation is within results from a given team. BRW’s headline finding conflates within- and between-team variation. That results vary across specifications is routine, and routinely reported. In fact, results in the original study (4) span a similarly broad range (Figure 1). It is therefore contestable that variation “remains hidden when considering a single study in isolation” (1).

The data are inadequate for the hypothesis. Given the causal nature of the hypothesis and observational nature of the data, there are inevitable limitations to the conclusions that can be drawn. Moreover, the inherent difficulty of measuring human attitudes is often underappreciated. One team “conducted preliminary measurement scaling tests, concluded that the hypothesis could not be reliably tested, and thus, did not design or carry out any further tests” (1). With more suitable data, one wonders if the results had not shown greater convergence, or at least more meaningful variation.

Limitations notwithstanding, BRW’s study is a landmark in crowdsourced open science. As the authors note, the underspecified nature of hypotheses and identification is representative of much published work. The questions raised in this comment do not diminish the

challenges social scientists face in making their discipline more credible. Researcher degrees of freedom remain an important and underrecognized source of uncertainty. Recognizing it should not, however, detract from underlying challenges that may turn out to matter as much if not more: theory, measurement, and causal inference (5–10).

## References

1. N. Breznau, et al., Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proc. Natl. Acad. Sci. U.S.A.*, 119(44), e2203150119 (2022).
2. R. Silberzahn, et al., Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Adv. Meth. Pract. Psychol. Sci.*, 1(3), 337356 (2018).
3. N. Huntington-Klein, et al., The influence of hidden researcher decisions in applied microeconomics. *Econ. Inq.*, 59(3), 944-960 (2021).
4. D. Brady, R. Finnigan, Does immigration undermine public support for social policy? *Am. Sociol. Rev.*, 79(1), 17-42 (2014).
5. K. Auspurg, J. Brüderl, Has the credibility of the social sciences been credibly destroyed? Reanalyzing the “Many Analysts, One Data Set” project. *Socius*, 7, 23780231211024421 (2021).
6. P. Engzell, J. M. Rohrer, Improving social science: Lessons from the open science movement. *PS Polit. Sci. Polit.*, 54(2), 297-300 (2021).
7. J. K. Flake, E. I. Fried, Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Adv. Meth. Pract. Psychol. Sci.*, 3(4), 456-465 (2020).
8. I. Lundberg, et al., What is your estimand? Defining the target quantity connects statistical evidence to theory. *Am. Sociol. Rev.*, 86(3), 532-565 (2021).
9. M. Muthukrishna, J. Henrich, A problem in theory. *Nat. Hum. Beh.*, 3(3), 221- 229 (2019).
10. J. M. Rohrer, Thinking clearly about correlations and causation: Graphical causal models for observational data. *Adv. Meth. Pract. Psychol. Sci.*, 1(1), 27-42