



# DEMOGRAPHIC RESEARCH

*A peer-reviewed, open-access journal of population sciences*

---

## ***DEMOGRAPHIC RESEARCH***

**VOLUME 41, ARTICLE 32, PAGES 949–952**

**PUBLISHED 10 OCTOBER 2019**

<https://www.demographic-research.org/Volumes/Vol41/32/>

DOI: 10.4054/DemRes.2019.41.32

*Editorial*

**Editorial:**

**P-values, theory, replicability, and rigour**

**Jakub Bijak**

© 2019 Jakub Bijak.

*This open-access work is published under the terms of the Creative Commons Attribution 3.0 Germany (CC BY 3.0 DE), which permits use, reproduction, and distribution in any medium, provided the original author(s) and source are given credit.*

*See <https://creativecommons.org/licenses/by/3.0/de/legalcode>.*

## Contents

1	Introduction	949
2	P-values, theory, replicability, and rigour: Guidance for authors	950
	References	952

## **Editorial: P-values, theory, replicability, and rigour**

**Jakub Bijak,<sup>1</sup> on behalf of the Editorial Board**

### **Abstract**

#### **BACKGROUND**

In the light of the recent discussions about the statistical rigour of empirical research, including the interpretation and use of p-values and the importance of the theoretical underpinnings of population studies, the editorial board of *Demographic Research* has adopted dedicated guidance for authors. Its aim is to clarify our expectations and highlight good practice in these areas. Starting from Volume 42 (2020), authors will be encouraged to follow these guidelines.

### **1. Introduction**

The last few years have witnessed a lively discussion questioning the rigour of a lot of empirical research that relies on the application of statistical methods, especially in the social sciences. Doubts have spread about the reproducibility and replicability of a large number of literature findings, first in psychology (Open Science Collaboration 2015), and then more broadly. One of the main concerns, already noted some time ago (e.g., Gigerenzer 2004), is that a misuse or misinterpretation of statistical methods can lead to “false positives, overhyped claims and overlooked effects” (*Nature* Editorial 2019: 283), creating publication bias against ‘non-significant’ or negative results (for an early warning, see Sterling 1959 and, recently, Fanelli 2012).

Concerned about the misuse of some commonly used tools, the American Statistical Association has proposed a set of guidelines on the use, interpretation, and limitations of ubiquitous statistical measures such as p-values (Wasserstein and Lazar 2016; Wasserstein, Schirm, and Lazar 2019). These guidelines are very much in line with what was suggested to the *Demographic Research* community by Jan Hoem over a decade ago (Hoem 2008). A large part of applied demographic work – although by no

---

<sup>1</sup> Editor, *Demographic Research*; University of Southampton, UK. Email: [J.Bijak@soton.ac.uk](mailto:J.Bijak@soton.ac.uk).

means all – uses statistical methods in some form, even if population sciences take pride in utilising a far broader range of tools, including those that are specific to demography, such as projections or indirect methods. At the same time, some other statistical tools are underutilised; for example, randomised control trials, and, to a lesser extent, observational studies. What is also currently lacking in our discipline is thorough theoretical reflection (Burch 2003, 2018).

Given that all these developments may have a profound impact on the ways applied social science is carried out in the future, and thus on the journals that rely on the use of statistics, the editorial board of *Demographic Research* has adopted guidance for authors, which follows below.

## **2. P-values, theory, replicability, and rigour: Guidance for authors**

The aim of this guidance is to clarify our expectations and highlight good practice in empirical research that we seek to publish. By so doing we want to ensure the statistical – or, more broadly, analytical – rigour of empirical papers that we publish in *Demographic Research*, while not potentially discouraging any applied work which makes important contributions to the field. The guidance consists of seven points, inspired by the past reflections in our journal (Hoem 2008), and the recent discussion in the statistical community (Wasserstein, Schirm, and Lazar 2019), and includes specific areas that are important from the point of view of demography and population sciences.

1. Articles in *Demographic Research* should focus on rigorous and innovative thinking and on making meaningful contributions to demographic debates. We welcome positive and negative results alike, as long as they tell us something important about the phenomenon at hand.
2. We encourage submissions in the category Descriptive Findings, which has been designed for early exploratory and descriptive papers. The Research Articles category is reserved for more advanced exploratory work and for confirmatory and methodological papers, while Research Materials for pedagogical and practical articles.
3. We also encourage theoretical reflection in a wider sense, including a broader use of formal models as tools of theoretical description where appropriate (Burch 2003 and 2018).
4. We discourage the use of the phrase ‘statistically significant’ or (in the same context) ‘significant’, as well as the use of asterisks and similar symbols referring to discretised p-value intervals. The conclusions of papers should not be justified by the magnitude of p-values alone (Hoem 2008; Wasserstein and Lazar 2016).

5. We encourage authors to reflect unflinchingly on the quality of the data they use and expect frank acknowledgement of the various imperfections that characterize their data sources. Analyses that extend beyond the generic caveats of accuracy and representativeness to provide more thorough engagement with the consequences of sampling and non-sampling errors are especially welcome.
6. We ask authors to be open about the uncertainty of results. Reporting of credible intervals, confidence intervals, or similar uncertainty measures alongside the effect sizes is preferable to p-values (see also Hoem 2008). If p-values are reported they should be presented as continuous measures, and not discretised. Uncertainty can be reported in many different ways: as confidence or credible intervals or as error bars or bands in the charts, as long as what is represented is clear to the readers.
7. We encourage testing different specifications of models, as well as sensitivity to key assumptions and analytical choices, before making specific conclusions (see Steegen et al. 2016). This is especially important for exploratory studies, although of course some form of robustness check remains good practice for other submissions as well.

Finally, reaffirming the journal's Open Science commitments, in place since 2013, we strongly encourage adherence to replicability principles. Henceforward, not only can Replicable status for a paper be earned by providing programme code and (where possible) data or meta-data, but we are also inviting replications of previously published studies, possibly in different contexts. To this end, we are pleased to introduce another article category, Replication, with the following description:

*Replications are carefully prepared and executed studies aimed at replicating other results on the same topic, published either in Demographic Research or elsewhere in the literature. The replication can be carried out in the same context as the original work or in a different context, and the results should illuminate and reflect the similarities to and differences from the original study.*

In practical terms, starting from submissions for Volume 42 (2020), we ask authors to read and acknowledge the contents of this guidance, and to follow it wherever possible. Where appropriate, the editors responsible for individual papers will request reasonable amendments to submitted papers in order to align them more closely with the guidance. We strongly believe that encouraging good practice in the areas covered by this guidance will be conducive to producing better, more robust, and more original science, and will help strengthen the research contributions of applied demographic work. We will be working with authors to achieve this aim.

## References

- Burch, T.K. (2003). Demography in a new key: A theory of population theory. *Demographic Research* 9(11): 263–284. doi:10.4054/DemRes.2003.9.11.
- Burch, T.K. (2018). *Model-based demography: Essays on integrating data, technique and theory*. Cham: Springer. doi:10.1007/978-3-319-65433-1.
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics* 90(3): 891–904. doi:10.1007/s11192-011-0494-7.
- Gigerenzer, G. (2004). Mindless statistics. *Journal of Socio-Economics*: 33(5): 587–606. doi:10.1016/j.socec.2004.09.033.
- Hoem, J.M (2008). The reporting of statistical significance in scientific journals: A reflexion. *Demographic Research* 18(15): 437–442. doi:10.4054/DemRes.2008.18.15.
- Nature Editorial (2019). Significant debate: It’s time to talk about ditching statistical significance. *Nature* 567: 283. doi:10.1038/d41586-019-00874-8.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349(6251): aac4716.1-8. doi:10.1126/science.aac4716.
- Steen, S., Tuerlinckx, F., Gelman, A., and Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science* 11(5): 702–712. doi:10.1177/1745691616658637.
- Sterling, T.D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *Journal of the American Statistical Association* 54(285): 30–34. doi:10.1080/01621459.1959.10501497.
- Wasserstein, R.L. and Lazar, N.A. (2016). The ASA’s statement on p-values: Context, process, and purpose. *The American Statistician* 70(2): 129–133. doi:10.1080/00031305.2016.1154108.
- Wasserstein, R.L., Schirm, A.L., and Lazar, N.A. (2019). Moving to a world beyond ‘ $p < 0.05$ ’. *The American Statistician* 73(Suppl. 1): 1–19. doi:10.1080/00031305.2019.1583913.