DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 50, ARTICLE 43, PAGES 1265–1280
PUBLISHED 7 JUNE 2024
https://www.demographic-research.org/Volumes/Vol50/43/
DOI: 10.4054/DemRes.2024.50.43

Reflection

Open science practices in demographic research: An appraisal

Ugofilippo Basellini

Contents

Demographic Research: Volume 50, Article 43
Reflection

Open science practices in demographic research: An appraisal

Ugofilippo Basellini[1]

Abstract

BACKGROUND
In the light of recent concerns about the reliability of scientific research, the open science movement has attracted considerable attention and interest from a variety of sources, including researchers, research institutions, the business sector, intergovernmental organisations, the media, and the public. However, the current extent of openness in demographic research remains unknown.

METHODS
All relevant publications in four leading journals of anglophone demography – *Demography*, *Population and Development Review*, *Population Studies*, and *Demographic Research* – over the last decade are analysed. Using a text-search algorithm, two quantitative metrics of open scientific knowledge are estimated: the share of publications that can be openly accessed, and the share of publications providing open software codes for reproducibility or replicability purposes.

RESULTS
Two contrasting patterns emerge from these indicators. Access to demographic research papers is increasingly available to everyone, with more than 90% of open-access publications in 2023. Conversely, the provision of open software codes has been and still remains considerably low, with only small signs of improvement over time. Over the last three years, on average 31% of articles in *Demographic Research* provided these materials and only about 12% in the other journals.

CONTRIBUTION
This reflection provides the first assessment of the adoption of some open science practices in demographic research and their evolution over the last decade. An urgent change is needed in the sharing of software codes (along with the data used, where possible) to contribute to the advancement of demographic research. Some recommendations for promoting this change are discussed.

[1] Max Planck Institute for Demographic Research, Rostock, Germany. Email: basellini@demogr.mpg.de.

1. Introduction

Sharing is caring. This phrase captures the spirit of recent and growing concerted efforts to promote open science across scientific disciplines. But what exactly is open science? Since its first mentions in the early 2010s, open science has lacked a formal definition, with varying interpretations and levels of awareness in different fields (European Commission 2015; Vicente-Saez and Martinez-Fuentes 2018). For some, open science merely refers to open-access scientific publications; for others, it signifies the open provision of data and software codes; and for still others, it concerns the participation of non scientists into the research process under the label of citizen science (Mirowski 2018). Acknowledging the need for a common definition of open science with a shared set of values and principles, the UNESCO issued a Recommendation on Open Science in 2021, which was unanimously adopted by its 193 member states. The recommendation provides the first internationally agreed definition of open science, consisting in a set of principles and practices that aim to make "multilingual scientific knowledge [...] accessible and reusable for everyone [...] for the benefits of science and society" and that are based on four key pillars: open scientific knowledge, open science infrastructures, open engagement of societal actors, and open dialogue with other knowledge systems (United Nations Educational, Scientific and Cultural Organization 2021).

Several factors contributed to the growth of the open science movement during the last decade. Although concerns had already been raised about the validity of several research findings across disciplines (Ioannidis 2005), it was not until the early 2010s that the field of psychology was shaken by a replication crisis: A series of papers uncovered the pervasiveness of questionable research practices as well as failures to replicate published findings in the discipline (e.g., Bem 2011; Simmons, Nelson, and Simonsohn 2011; Open Science Collaboration 2015). At the same time, several high-profile fraud cases involving data fabrication and plagiarism have emerged in multiple disciplines (for some examples, see, Christensen, Freese, and Miguel 2019). Furthermore, retractions of scientific papers have increased over time (Van Noorden 2023), along with well-publicised cases of companies that produce and sell fraudulent manuscripts that resemble genuine research (so-called paper mills) (Else and Van Noorden 2021).

In addition to addressing these problems, open science has received considerable interest for the benefits that it aims to bring, for example, in terms of scientific knowledge. Open-access scientific articles allow scientific research and knowledge to spread and circulate much wider than if published behind a paywall. Open research data and open software codes provide the means to verify, generalise, and build upon research results, while also providing a social control mechanism that discourages fraudulent work (Freese and Peterson 2017). Questionable research practices, such as HARKing (Hypothesizing After the Results are Known) (Kerr 1998) or p-hacking (selectively choosing data and analysis that lead to statistically significant results, as measured by p-values) (See,

e.g., Wasserstein and Lazar 2016), can be transparently avoided by pre-registering research hypotheses, methods, and analysis prior to conducting a scientific study. All these practices contribute to increasing confidence in the successful functioning of science.

Reproducibility and replicability are two key concepts in open science as they play a crucial role in assessing the credibility of research results and, more generally, of scientific research. These concepts are defined and implemented differently across scientific disciplines (Freese and Peterson 2017; Rahal et al. 2022). In this article, the focus will be on reproducibility and replicability of empirical studies, using the definitions provided by the National Academies of Sciences, Engineering, and Medicine (2019) and Nosek et al. (2022). Specifically, reproducibility refers to testing the reliability of a prior finding using the same data and the same analysis strategy, while replicability refers to testing the reliability of a prior finding with different data. Thus, the public sharing of research data and, more importantly, of software codes is a fundamental feature for enabling and facilitating reproduction or replication attempts.

Despite significant progress towards open science in recent years, science today is still not fully open. Most scientific articles can be accessed only by paying a subscription fee (National Academies of Sciences, Engineering, and Medicine 2018) and can be made open access only by paying expensive article processing fees. Sharing of open data and software codes is becoming more common but is still not routine across all disciplines. In addition, there are other significant barriers to open science, such as lack of infrastructures, incentives, and training for researchers, as well as privacy, security, and proprietary barriers to data sharing (National Academies of Sciences, Engineering, and Medicine 2018). Moreover, scientific disciplines differ in the nature and methods of research, making it difficult to define universal norms of open science.

Demography – the study of population processes – is a social science discipline that is theoretically well suited to the implementation of open science practices, particularly in terms of sharing data and software codes. Most of our work, even when based on formal equations or sociological theories, is empirical in practice, often aspiring to guide governments and policymakers. Concepts such as reproducibility and replicability should be familiar and widely embraced by the field. However, the degree to which demographic research is open is currently unknown. The aim of this reflection is to examine how open science practices related to scientific knowledge have been adopted within the discipline of demography over the last decade. Specifically, looking at four leading journals of anglophone demography, two quantitative metrics of open scientific knowledge are analysed: the share of open-access publications in the field, and the share of articles that provide access to open software codes for reproducibility or replicability purposes. The potential reasons behind the observed trends are discussed, along with some recommendations for increasing the sharing of software codes (along with the data used, where possible) of published demographic research.

2. Open scientific knowledge in demographic research

All publications during the period 2013–2023 in *Demographic Research*, *Demography*, *Population and Development Review*, and *Population Studies* were downloaded in January 2024. Of these publications, research articles and research notes – typically based on empirical data in demography – were retained, while book reviews, corrections, and editor notes were excluded from the analysis. These journals were selected because they are largely considered to be the four leading journals of anglophone demography, covering the full range of demographic topics rather than specialising in a single subject. The first year of analysis corresponds to the establishment of the 'Replicable' category of papers in *Demographic Research*, which also broadly coincides with the first mentions of open science in the early 2010s. Access to all articles was provided by the Max Planck Digital Library.

These publications were examined using a text-search algorithm to determine whether each publication was open access and whether open software codes were made available by the authors for reproducibility or replicability purposes. In practice, the text of each publication was imported into the statistical software `R` and searched for mentions of open-access publication and mentions of open codes for reproducibility or replicability. This analysis was performed using the `pdftools` package in `R` (Ooms 2023; R Core Team 2023).

Detecting whether an article is open access or not is relatively straightforward as this information is always provided in a standard format on the first or last page of the article, along with the associated licence. Conversely, detecting the provision of open software codes is more challenging as there is no standard way of reporting this information, and the wording used varies considerably between authors. Therefore, a two-step procedure was used to identify the provision of software codes. First, a liberal initial screening approach was used to minimise the chances of not identifying articles that provided this material. More than 600 keywords, consisting of different combinations of adjectives, nouns, and verbs typically used to refer to open software codes, along with popular repositories of data and codes (e.g., GitHub, Open Science Framework, Zenodo, Dataverse, and Figshare), were used for this purpose. Two often-encountered examples of these keywords are "replication files" and "code is available"; the full list of keywords can be found in the replication package of this article (see Section 4, Data availability statement). Second, a manual screening of all identified instances of open software codes was performed to remove all papers that did not provide such materials. During this screening, all instances of codes to be made available at a future date or "upon request" were also excluded since these latter requests often go unfulfilled (see, e.g., Tedersoo et al. 2021; Krähmer, Schächtele, and Schneck 2023). A similar approach, based on 5 keywords and without a second-stage verification, is employed by Alexander (2022) to detect the availability of either data or code, or both, in *Demography* papers since 2011.

For the journal *Demographic Research*, it was also possible to obtain more precise estimates of the number of articles that provide open software directly from the journal's website. Since 2013, the journal has a dedicated 'Replicable' category for any publication providing open study materials. Specifically, a publication is marked as Replicable "if authors provide data and programs for reproducing calculations".[2] As such, for each published volume, the total number of publications as well as the number of Replicable papers was extracted from the journal website. Furthermore, in order to investigate whether the introduction of this category contributed to increasing the sharing of open software codes, the text-search analysis for this journal was extended to the years 2010–2012.

To test the accuracy of the two-stage text-search algorithm, a manual inspection of two random issues of *Demography*, *Population and Development Review*, and *Population Studies*, and of one random volume of *Demographic Research* was performed to verify that open access and open codes were indeed correctly identified. In the corresponding 113 articles, there were no misclassifications of either metric. However, 7 articles that were marked as Replicable on the *Demographic Research* website were not identified as such by the text search because the provision of codes as Additional Files of the publication was not explicitly stated in the paper (and hence not detectable by the text search). This issue was also confirmed by manual inspection of other Replicable articles in the journal. The reason for this could be related to the fact that *Demographic Research* is an online-only journal, that Replicable papers are clearly marked on the journal website with a replication badge, and that all such papers are grouped together in a separate collection accessible through the website. These factors may have prevented some authors from including relevant information directly in the text.

Table 1 shows, for each of the four journals, the volumes and issues included in the analysis, along with the total number of relevant publications, the estimated number and proportion of publications identified as open access and with open software codes, and whether these estimates were derived from the text search or the journal's website. The table clearly shows that a considerably small number of publications adhere to the practices of open scientific knowledge. *Demographic Research* outperforms the other journals, with all of its publications being open access, and with 28.9% publications providing open codes (as per the journal's website estimates). For the other three journals, the shares of both indicators are considerably lower, with about 33% to 38% open-access publications, and only 4% to 7% publications providing open software codes. The discrepancy in the open software estimates for *Demographic Research* from the two sources is mainly due to the non-reporting issue mentioned above, and as such, the text-search estimates for this journal should be treated with caution. However, it is worth noting here that this particular bias should not affect other journals, which do not have a Replicable category. Regarding the text-search algorithm for the four journals, the first stage of the

[2] See https://www.demographic-research.org/articles/replicablearticles.

analysis initially identified 442 papers as providing open software, and the second-stage manual verification of these publications reduced this number to 200.

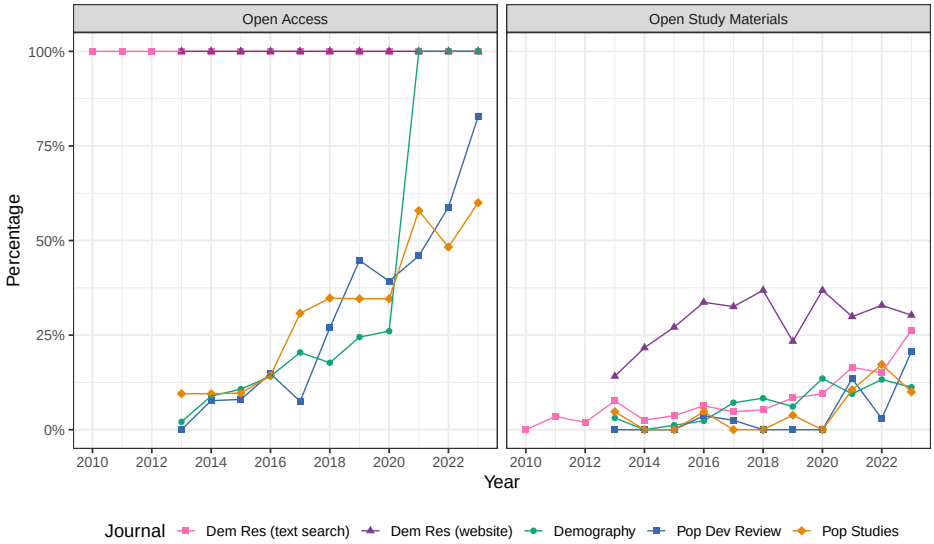
Table 1: Volumes and issues of the four journals included in the analysis, together with the total number of relevant publications, the estimated number (proportion in brackets) of publications identified as open access and with open software codes, and the source of these estimates

Journal	Volume (issue)	Publications	Open access (%)	Open software (%)	Source
Demographic Research	28–49	1,102	1,102 (100%)	319 (28.9%)	Website
Demographic Research	22–49	1,283	1,283 (100%)	100 (7.8%)	Text search
Demography	50(1)–60(6)	1,027	391 (38.1%)	71 (6.9%)	Text search
Population and Development Review	39(3)–49(4)	324	107 (33%)	14 (4.3%)	Text search
Population Studies	67(1)–77(3)	292	98 (33.6%)	15 (5.1%)	Text search

Notes: The first two issues of volume 39 of *Population and Development Review* were not included due to errors in reading the PDF files of the articles in `R` using `pdftools`.

Figure 1 provides a more nuanced account on the adoption of these open science practices in the four journals over time. The two panels of Figure 1 highlight two different rates of progress in the two indicators of open knowledge analysed. On the one hand, the proportion of open-access publications has been constantly growing over time. *Demographic Research* has led the way in this respect, making all papers openly available since its foundation in 1999. *Demography* joined this philosophy in 2021, when it became a diamond open-access journal with only open-access publications, which were only about 25% in 2020. The proportion of openly available publications in *Population and Development Review* and *Population Studies* has been constantly increasing over time, especially for the former journal since 2017. In 2023, approximately 92% of all publications across the four journals were open access.

Figure 1: Estimated percentage of published articles that are open access (left panel) and that have open software codes (right panel) in four leading journals of anglophone demography, 2010–2023



On the other hand, the right panel of Figure 1 shows that the proportion of articles providing open software codes was rather low in all years analysed, again with *Demographic Research* outperforming other journals. In this journal, the indicator shows an increasing trend in the first few years after the introduction of the Replicable category, but it starts to fluctuate from 2018 onwards. For the other journals, it is possible to observe some fluctuations in the first years of the analysis and some small signs of improvement in more recent years, with an average provision of about 12% in the last three years (against a 31% in *Demographic Research*).

To investigate whether the introduction of the Replicable category in *Demographic Research* contributed to an increase in the proportion of articles with open software codes, the text-search analysis of the journal was extended to the years 2010–2012, also shown in Figure 1. The rather constant proportion of open codes detected by the text-search algorithm, especially before and after 2013, contrasts with the increasing trend observed from the journal's website estimates. However, the text-search estimates for this journal are not as accurate for the reasons mentioned above, so they may not be very helpful in answering this question (see also Section 3, Discussion).

3. Discussion

The goal of this reflection has been to assess the adoption of open science practices in demographic research over the last decade. Two contrasting trends emerged from analysing all relevant publications of four leading journals of anglophone demography: While research papers are increasingly accessible to all, with open-access publications becoming the norm rather than the exception, the sharing of software code for reproducibility or replicability purposes is still very limited, with only faint signs of improvement over the last three years.

As per the 2021 UNESCO recommendation, open science is a broad movement with different facets. In this reflection, only open scientific knowledge was analysed, considering specifically four anglophone demographic journals. Clearly, these findings do not generalise to the broader demographic discipline, as demographic research is also published in other languages and in other demographic or neighbouring disciplines journals (for a recent overview of demography's interdisciplinary reach, see Merli et al. 2023). Nonetheless, the results provide a first appraisal and important insights on recent and current open science practices within the field.

The two metrics of open scientific knowledge analysed here are closely related to each other since reproducibility or replicability of published findings cannot be actually achieved unless the publication is accessible to everyone. In this sense, it is encouraging to observe a rapid transition to a very high share of open-access publications. This speeds up the process of disseminating research, building on results, and allowing wider and more inclusive participation in research (National Academies of Sciences, Engineering, and Medicine 2018). There is also evidence that open-access articles receive more citations and media coverage than non-open-access articles (McKiernan et al. 2016; Wang et al. 2015). However, concerns remain regarding who can publish openly. *Population and Development Review* and *Population Studies*, among several other scientific journals, require the payment of expensive article processing charges (APCs) for publishing open access. While agreements exist between publishing companies and some institutions and even countries to cover or reduce these costs, it is often the case that researchers from less-advantaged countries, or from less-funded institutes and universities within a country, do not receive sufficient financial support to publish openly. *Demographic Research* has been an admirable example of open-access publishing, and it is desirable that other journals will follow *Demography*'s transition to diamond open access[3] to reduce these global inequalities. A current alternative available to everyone is to publish preprints or working papers, either prior to journal peer review or as accepted manuscripts prior to copyediting – in institutional archives, preprint servers, or personal websites (Sherpa Romeo provides a comprehensive review of journal open-access policies across disci-

[3] It should be noted that, in 2024, the journal moved to a "Subscribe to Open" model, according to which articles will be made open access only if a funding threshold is met (which has not happened as of April 2024).

plines[4]). This practice contributes to the openness of research beyond what is captured by the text-search algorithm used here.

One of the reasons why science is still not fully open relates to the lack of incentives to make such transition. The publishing industry has strong incentives to keep articles behind paywalls to continue profiting from high APCs and typically unremunerated authors and reviewers. The 'publish or perish' pressure faced by a constant number of scientists facing an exponential growth of scientific publications (Hanson et al. 2023), along with journals' interest in publishing almost exclusively positive findings (so-called publication bias) (See, e.g., Dickersin 1990), can incentivise questionable research practices. More importantly, there currently lacks clear incentives for scientists to share their study materials since, in the vast majority of cases, citation counts and number of publications are the most used indicators of a scientist's productivity, discouraging time investments in learning, teaching, and adopting open science practices. This is particularly relevant for early career researcher and especially PhD students, who are often required to publish a number of articles in a limited amount of time to obtain their degree, and researchers from groups discriminated against in academia, such as women and non-white scholars. Changing the structure of incentives in the academic system – for example, by placing more value on reproducible or replicable studies – would be an important step towards improving the current scientific landscape.

The interpretations of reproducibility and replicability of empirical studies used in this article were rather strict, meaning that the direct availability of software codes was considered essential for reproducibility or replicability purposes. Less strict interpretations of the concepts are also possible: For example, one could try to replicate some published results by independent coding, closely following the description of the methods used by the authors. Clearly, such an approach presupposes a great deal of familiarity with both the methods and the data, which may not always be available to users, as well as a considerable amount of time, the lack of which may discourage such attempts in the first place. Although demographers have always taken great care to carefully describe the data and methods they have used and developed, the public provision of software codes goes a long way towards helping others by removing barriers to reproducibility and replicability and by avoiding the duplication of costs associated with redeveloping codes that are not openly shared.

Following the same line of reasoning, open software code was considered an essential element of the text search, whereas data availability was not considered as such. Indeed, open software code is needed for both reproducibility and replicability (in their strict interpretations). Demographers have long been at the forefront of making data available to the public, with several examples of large datasets made available for use by anyone for several decades (such as, for example, the Human Mortality Database, Human

[4] See https://www.sherpa.ac.uk/romeo/.

Fertility Database, Integrated Public Use Microdata Series (IPUMS), and Demographic and Health Surveys databases, the various datasets provided by government agencies, etc.). Also for this reason, the text search was based only on the availability of software codes. Nevertheless, it is worth keeping in mind that the provision of software code is not a panacea and does not guarantee reproducibility or replicability since incomplete data, missing instructions, software updates, and lack of information can result in considerable hurdles and even inability to reproduce results (Prike 2022). As such, providing a clear workflow, metadata and documentation can be as important as providing codes and data.

The text-search algorithm employed here for the detection of open software codes has some limitations that should be considered when interpreting the results. First, it is hard to quantify the under-detection of study materials and the related uncertainty of the estimates. A manual inspection of some random issues and volumes of the four journals was employed to increase the confidence of the text search; while this analysis increases confidence in the accuracy of the text-search algorithm, it does not eliminate the possibility of under-detecting the availability of open software codes. Second, the text search is not powerful enough to differentiate between instances of reproducibility and replicability. Future research could further investigate differences in the uptake of these two concepts, as well as improve the text-search algorithm using, for instance, machine-learning techniques. Third, only one person (the author of this paper) performed the manual inspection of the articles, which may have contributed to some misclassification of papers.

The text-search inspection of *Demographic Research* articles did not provide a convincing argument for the introduction of the Replicable category. However, the remarkably higher proportion of articles providing open software codes compared to the other three journals, as well as the initial upward trend of Replicable papers derived from the journal's website, seem to suggest that this strategy has contributed to a greater willingness to share codes and data. The clear credit given to authors investing on this – via the replication badge of the published article – may have been an important incentive for this change. Indeed, the effectiveness of open science badges for increasing data and code sharing has already been documented in the literature (Kidwell et al. 2016), although these badges do not necessarily guarantee full reproducibility or replicability of a paper's findings (Crüwell et al. 2023; Hardwicke et al. 2021). Nevertheless, the proportion of articles with open software codes is very low in the four journals analysed, with only small signs of improvement over the last three years. This is detrimental to our scientific discipline since the lack of study materials makes it much more difficult, if not impossible, to reproduce, replicate, and advance demographic research. Although not all demographic papers should be reproducible and replicable in the empirical sense considered here – for example, a minority of papers published in these four journals are review or policy articles that do not engage with empirical analysis – it is evident that the current proportion of papers with open software codes is highly unsatisfactory. In an effort to

reaffirm the journal's open science commitments, *Demographic Research* introduced in 2019 a new category of articles, denoted 'Replication', inviting "carefully prepared and executed studies aimed at replicating other results on the same topic, published either in *Demographic Research* or elsewhere in the literature" (Bijak 2019). Sadly, only four Replication articles have been published as of April 2024, highlighting how little these studies are valued and performed in demography.

There is an urgent need for increasing the sharing of data and codes of published demographic research articles. While there could be good reasons for not sharing data (e.g., privacy or proprietary nature of individual-level data), nothing should prevent researchers, in principle, from sharing software codes, which could be fruitfully employed by others on different datasets for replicability purposes or for new studies. Moreover, even when individual-level data cannot be openly shared, a provision of the derived aggregate-level data or simulated individual-level dataset upon which running the codes could still be of benefit. Open-access repositories, such as Zenodo, the Open Science Framework, and GitHub, offer practical ways to share codes and data. Sharing of research materials is something that researchers can directly control. Choosing to do so not only contributes to the advancement of science but also benefits the authors themselves: Articles with available data and codes receive more citations than similar studies without available data and codes (Piwowar and Vision 2013; Vandewalle 2012). Moreover, journals could encourage authors to deposit their materials in repositories that provide a digital object identifier (such as Zenodo or the Open Science Framework), which could be listed in the references, helping authors to gain citations.

From an institutional perspective, an open science culture could be encouraged by, for example, providing regular courses, workshops, and training on open science practices. Some researchers, and especially young scholars, might not share their codes for a lack of confidence in their work. Building an institutional or even discipline-wide culture of support for sharing would alleviate such fears. Recognition of open science efforts – for example, with dedicated awards (for both publishing and teaching) – would also further incentivise efforts in this direction. Finally, it is noteworthy that an increasing number of journals require a 'Data Availability Statement' at the time of manuscript submission. Of the journals analysed here, only *Population and Development Review* currently has this requirement – which, however, does not appear to have been enforced in recent years – and this may be a top-down editorial strategy to increase reproducibility and replicability in demographic research. An even stricter editorial strategy would be to require a replication package – as recently enforced by the journal *Sociological Science* – as a condition of publication for articles that rely on statistical or computational methods. Indeed, recent evidence suggests that such policies can successfully increase reproducibility (Fišar et al. 2024).

4. Data availability statement

All results for the open-access publications are fully reproducible using the data and source codes available at the OSF repository https://osf.io/3gdzc/. The repository also contains the text-search results for all 2,926 papers analysed in this article (in the file "03-combined-cleaned.Rdata").

5. Acknowledgements

The author is an Associate Editor for the journal *Demographic Research*, but this paper was not commissioned by the journal or its Editorial Board, and it was subject to the standard peer-review process. The author would like to thank the Max Planck PhDnet Open Science Group and the Max Planck Digital Library for organising the Open Science Ambassadors Meeting in September 2023, and the other participants for the fruitful discussions on open science that inspired the writing of this article. The author further thanks Aliakbar Akbaritabar, Diego Alburez-Gutierrez, Esther Denecke, Ana Cristina Gomez Ugarte, Ilya Kashnitsky, Marília Nepomuceno, Rima-Maria Rahal, Jonas Schöley, and Emilio Zagheni for useful comments on a previous version of this article, and Tom Theile for assistance on the text search. Access to non-open-access publications was provided by the Max Planck Digital Library.

References

Alexander, M. (2022). Reproducibility in demography. Presentation given at the 2022 Toronto Workshop on Reproducibility. Slides and codes available at https://github.com/MJAlexander/reproducibility-in-demography/tree/master.

Bem, D.J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology* 100(3): 407. doi:10.1037/a0021524.

Bijak, J. (2019). P-values, theory, replicability, and rigour. *Demographic Research* 41(32): 949–952. doi:10.4054/DemRes.2019.41.32.

Christensen, G., Freese, J., and Miguel, E. (2019). *Transparent and reproducible social science research: How to do open science*. Berkeley, CA: University of California Press.

Crüwell, S., Apthorp, D., Baker, B.J., Colling, L., Elson, M., Geiger, S.J., Lobentanzer, S., Monéger, J., Patterson, A., Schwarzkopf, D.S., Zaneva, M., and Brown, N.J.L. (2023). What's in a badge? A computational reproducibility investigation of the open data badge policy in one issue of Psychological Science. *Psychological Science* 34(4): 512–522. doi:10.1177/09567976221140828.

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *JAMA: The Journal of the American Medical Association* 263(10): 1385. doi:10.1001/jama.1990.03440100097014.

Else, H. and Van Noorden, R. (2021). The fight against fake-paper factories that churn out sham science. *Nature* 591(7851): 516–520. doi:10.1038/d41586-021-00733-5.

European Commission (2015). Open innovation, open science, open to the world – A vision for Europe. doi:10.2777/061652. Luxembourg: Publications Office of the European Union.

Fišar, M., Greiner, B., Huber, C., Katok, E., and Ozkes, A.I. (2024). Reproducibility in Management Science. *Management Science* 70(3): 1343–1356. doi:10.1287/mnsc.2023.03556.

Freese, J. and Peterson, D. (2017). Replication in social science. *Annual Review of Sociology* 43: 147–165. doi:10.1146/annurev-soc-060116-053450.

Hanson, M.A., Barreiro, P.G., Crosetto, P., and Brockington, D. (2023). The strain on scientific publishing. Cornell University (arXiv preprint 2309.15884). Available at https://arxiv.org/abs/2309.15884.

Hardwicke, T.E., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M.B., Peloquin, B.N., deMayo, B.E., Long, B., Yoon, E.J., and Frank, M.C. (2021). Analytic reproducibility in articles receiving open data badges at the journal Psychological Science: An observational study. *Royal Society Open Science* 8(1). doi:10.1098/rsos.201494.

Ioannidis, J.P. (2005). Why most published research findings are false. *PLoS Medicine* 2(8): e124. doi:10.1371/journal.pmed.0020124.

Kerr, N.L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review* 2(3): 196–217. doi:10.1207/s15327957pspr0203_4.

Kidwell, M.C., Lazarević, L.B., Baranski, E., Hardwicke, T.E., Piechowski, S., Falkenberg, L.S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T.M., Fiedler, S., and Nosek, B.A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology* 14(5): e1002456. doi:10.1371/journal.pbio.1002456.

Krähmer, D., Schächtele, L., and Schneck, A. (2023). Care to share? Experimental evidence on code sharing behavior in the social sciences. *PLoS ONE* 18(8): e0289380. doi:10.1371/journal.pone.0289380.

McKiernan, E.C., Bourne, P.E., Brown, C.T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B.A., Ram, K., Soderberg, C.K., Spies, J.R., Thaney, K., Updegrove, A., Woo, K.H., and Yarkoni, T. (2016). How open science helps researchers succeed. *eLife* 5. doi:10.7554/elife.16800.

Merli, M.G., Moody, J., Verdery, A., and Yacoub, M. (2023). Demography's changing intellectual landscape: A bibliometric analysis of the leading anglophone journals, 1950–2020. *Demography* 60(3): 865–890. doi:10.1215/00703370-10714127.

Mirowski, P. (2018). The future(s) of open science. *Social Studies of Science* 48(2): 171–203. doi:10.1177/0306312718772086.

National Academies of Sciences, Engineering, and Medicine (2018). *Open science by design: Realizing a vision for 21st century research*. Washington, DC: The National Academies Press. doi:10.17226/25116.

National Academies of Sciences, Engineering, and Medicine (2019). *Reproducibility and replicability in science*. Washington, DC: The National Academies Press. doi:10.17226/25303.

Nosek, B.A., Hardwicke, T.E., Moshontz, H., Allard, A., Corker, K.S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M.B., Rohrer, J.M., Romero, F., Scheel, A.M., Scherer, L.D., Schönbrodt, F.D., and Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology* 73: 719–748. doi:10.1146/annurev-psych-020821-114157.

Ooms, J. (2023). *pdftools: Text extraction, rendering and converting of PDF documents*. https://CRAN.R-project.org/package=pdftools. R package version 3.3.3.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349(6251): aac4716. doi:10.1126/science.aac4716.

Piwowar, H.A. and Vision, T.J. (2013). Data reuse and the open data citation advantage. *PeerJ* 1: e175. doi:10.7717/peerj.175.

Prike, T. (2022). Open science, replicability, and transparency in modelling. In: Bijak, J. (ed.). *Towards Bayesian model-based demography: Agency, complexity and uncertainty in migration studies. Methods Series, vol 17*. Cham: Springer International Publishing: 175–183. doi:10.1007/978-3-030-83039-7_10.

R Core Team (2023). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Rahal, R.M., Hamann, H., Brohmer, H., and Pethig, F. (2022). Sharing the recipe: Reproducibility and replicability in research across disciplines. *Research Ideas and Outcomes* 8. doi:10.3897/rio.8.e89980.

Simmons, J.P., Nelson, L.D., and Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22(11): 1359–1366. doi:10.1177/0956797611417632.

Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H. et al. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data* 8(1): 192. doi:10.1038/s41597-021-00981-0.

United Nations Educational, Scientific and Cultural Organization (2021). UNESCO recommendation on open science (SC-PCB-SPP/2021/OS/UROS). Available at https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en.

Van Noorden, R. (2023). More than 10,000 research papers were retracted in 2023: A new record. *Nature* 624(7992): 479–481. doi:10.1038/d41586-023-03974-8.

Vandewalle, P. (2012). Code sharing is associated with research impact in image processing. *Computing in Science and Engineering* 14(4): 42–47. doi:10.1109/mcse.2012.63.

Vicente-Saez, R. and Martinez-Fuentes, C. (2018). Open science now: A systematic literature review for an integrated definition. *Journal of Business Research* 88: 428–436. doi:10.1016/j.jbusres.2017.12.043.

Wang, X., Liu, C., Mao, W., and Fang, Z. (2015). The open access advantage considering citation, article usage and social media attention. *Scientometrics* 103(2): 555–564. doi:10.1007/s11192-015-1547-0.

Wasserstein, R.L. and Lazar, N.A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician* 70(2): 129–133. doi:10.1080/00031305.2016.1154108.