





Code sharing in ecology and evolution increases citation rates but remains uncommon

Brian Maitner^{1,2}  | Paul Efren Santos Andrade³  | Luna Lei⁴ | Jamie Kass⁵  |
 Hannah L. Owens^{6,7,8} | George C. G. Barbosa⁹ | Brad Boyle¹⁰ | Matiss Castorena¹⁰ |
 Brian J. Enquist^{10,11} | Xiao Feng¹² | Daniel S. Park^{13,14} | Andrea Paz^{15,16}  |
 Gonzalo Pinilla-Buitrago¹⁷ | Cory Merow¹⁸ | Adam Wilson²

Correspondence

Brian Maitner, Department of Integrative Biology, University of South Florida, St. Petersburg, Florida, USA.
 Email: bmaitner@gmail.com

Funding information

EU Horizon 2020 Marie Skłodowska-Curie grant, Grant/Award Number: 891702; Villum Fonden, Grant/Award Number: 25925; NASA Ecological Forecasting Team Applied Sciences Program, Grant/Award Number: 80NSSC21K1183; NASA BioSCape, Grant/Award Number: 80NSSC21K0086; US National Science Foundation, Grant/Award Number: 2225076 and 2225078

Abstract

Biologists increasingly rely on computer code to collect and analyze their data, reinforcing the importance of published code for transparency, reproducibility, training, and a basis for further work. Here, we conduct a literature review estimating temporal trends in code sharing in ecology and evolution publications since 2010, and test for an influence of code sharing on citation rate. We find that code is rarely published (only 6% of papers), with little improvement over time. We also found there may be incentives to publish code: Publications that share code have tended to be low-impact initially, but accumulate citations faster, compensating for this deficit. Studies that additionally meet other Open Science criteria, open-access publication, or data sharing, have still higher citation rates, with publications meeting all three criteria (code sharing, data sharing, and open access publication) tending to have the most citations and highest rate of citation accumulation.

KEYWORDS

code sharing, open access, open data, open science, R software, reproducibility

1 | INTRODUCTION

Reproducibility and transparency are cornerstones of reputable, rigorous, and mature science (Freedman et al., 2015; McNutt, 2014; Munafò et al., 2017; Nosek et al., 2015; “Reality check on reproducibility”, 2016). While definitions can vary by disciplines or change over time (Baker, 2016), reproducibility commonly refers to the ability to repeat previous analyses and obtain consistent results. Transparency refers to the degree or extent that details of the work, especially the methods and results, are

openly documented and communicated, thus allowing alternative interpretations (Drescher & Edwards, 2019; Wagenmakers et al., 2021). For programming scripts, the reproducibility spectrum begins with public, permanently archived code (Parker et al., 2016; Peng, 2011). In ecology and evolution, code (primarily in the statistical programming language R; Lai et al., 2019; R Core Team, 2023) has become the basis of most analyses (Feng et al., 2020), and the benefits of code sharing are increasingly recognized (Munafò et al., 2017; Parker, Forstmeier, et al., 2016). Clear, reusable code released under a permissive license (Stodden, 2009) ensures better

For affiliations refer to page 8.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Ecology and Evolution* published by John Wiley & Sons Ltd.

reproducibility of analyses, but it also may enhance the impact of publications (e.g., greater uptake of methods, more citations) and reduce duplicated efforts, allowing science to progress more effectively (McNutt, 2014; Munafò et al., 2017; Nosek et al., 2015; “Reality check on reproducibility”, 2016). Furthermore, well-documented code facilitates the peer review process, provides a valuable educational resource (Busjahn & Schulte, 2013), and facilitates our ability to credit developers, as software and package-usage data can be harvested directly from published code (Merow, Boyle, et al., 2023).

Has the increasing appreciation of code sharing influenced code sharing practices over time? Recent evidence suggests that biologists may be reluctant to share code. A study focused on publications in ecology journals with policies that mandated or encouraged code sharing found that 73% failed to share code (Culina et al., 2020), while a study focused on publications using agent-based models found that 81% did not provide code (Barton et al., 2022). PLOS open science indicators likewise suggest that code sharing is rare, with 92% of publications in Agricultural and Biological Sciences failing to share code (in comparison, 49% fail to share data; Public Library of Science, 2023). While some papers include the statement “code available upon request,” this promise is often not met (Stodden et al., 2018). Where published, code may also not be reusable due to licensing issues (Stodden, 2009). Resistance to code sharing and reuse may arise from unfamiliarity with best sharing practices, insecurity about code quality, fears of misuse or unsolicited appropriation of ideas, and excess preparation costs (Cadwallader & Hrynaszkiewicz, 2022; Gomes et al., 2022). However, it has been argued that many perceived issues with code sharing stem from misunderstandings of its risks and benefits (Gomes et al., 2022). To better understand how code sharing practices have changed over time and whether code sharing actually improves citation rates, we (1) estimated trends in R code sharing for articles in ecology and evolution published between 2010 and 2022 and (2) tested whether the citation rate was higher for papers that shared code. We focus on R because it has become the dominant coding language in ecology and evolution (Lai et al., 2019).

2 | MATERIALS AND METHODS

2.1 | Data collection

2.1.1 | List of ecology and evolution publications citing R

To generate a list of papers in ecology and evolution that likely made use of the R programming language (R Core Team, 2023), we performed a query on the Scopus database (<https://www.scopus.com>) using the *rscopus* R package (Muschelli, 2019). We searched Scopus (performed August 19, 2022) for peer reviewed journal articles that: (1) included the words “ecology” or “evolution” in an “all

fields” search (which searches article titles, keywords, abstracts, and journal titles); (2) were published in journals within the subject area “agriculture and biological sciences”; (3) were published after January 1, 2010; (4) were written in English (as this is currently the dominant language of publication in ecology and evolution; Mauranen et al., 2010); and (5) included a citation of R in their reference list.

2.1.2 | Checking for code and data availability

We manually evaluated a randomly chosen subset of the publications on our overall list. We selected a total of 1001 papers, evenly distributed across the time period (77 per year * 13 years). Papers that cited R but did not use it (or were unclear on whether they used it; $n=3$) were discarded and replaced by a randomly selected paper from the same year. For each publication in this subset, we manually identified whether the publication shared any R code, either as supplementary information, or via a link (e.g., to a Github repository). For each paper, we (i) checked for the presence of code in supplemental material, (ii) skimmed publications for code and data availability statements, (iii) searched through publications for terms associated with code (i.e., “code”, “supplement”, “appendix”, “R”, “script”, “Github”), and (iv) searched publications for URLs. Papers were scored with a binary variable indicating whether they shared R code or not. We did not distinguish between publications which shared sufficient code for reproduction and those which did not. We also did not attempt to rerun the code or assess its reproducibility, and only recorded the presence of any code, even if it was incomplete. Where code was included, we recorded the license the code was provided under, or lack thereof. We also assessed whether publications were open access and whether they shared open data in order to understand the importance of open code relative to these other open-access components. Open access information was provided by the *rscoopus* R package (Muschelli, 2019). Open data was scored as a binary variable indicating whether the authors shared the full set of raw data underlying the analyses or not. To control for differences in citation rates among journals, we downloaded impact factor information using the *scholar* R package (Keirstead, 2016) on June 16, 2023. To estimate the proportion of publications which use R but do not properly cite it, we screened 130 randomly selected publications evenly distributed across the time period. These publications were selected using identical criteria to the publications that did cite R, except that they did not include R in their list of references.

2.2 | Checking for code citations

Where code was shared in a citable location such as a DOI or URL ($n=33$), we assessed whether the code itself was cited by querying the Scopus database for the URL (and DOI, where appropriate) using

the *rscopus* R package (Muschelli, 2019). Publications where code was shared in appendices or supplementary information ($n=22$) were excluded, as there was no way of distinguishing citations of the code with citations of the publication itself.

2.3 | Analyses

All analyses were conducted in R version 4.3.0 (R Core Team, 2023). All R scripts underlying these analyses are available at: https://github.com/bmaitner/R_citations and via Zenodo (Maitner & Lei, 2024). For data processing, we used the R packages *stringdist* version 0.9.12 (van der Loo, 2014), *tidyverse* version 2.0.0 (Wickham et al., 2019), *googledrive* version 2.1.1 (D'Agostino McGowan & Bryan, 2023), and *googlesheets4* version 1.1.1 (Bryan, 2023); for analyses, the packages *bbmle* version 1.0.25.1 (Bolker & R Development Core Team, 2022), *DHARMA* version 0.4.6 (Hartig, 2022), *MuMIn* version 1.47.5 (Barton, 2023), *rscopus* version 0.6.6 (Muschelli, 2019), *rsq* version 2.6 (Zhang, 2018), *scholar* version 0.2.4 (Keirstead, 2016), and *stats* version 4.3.0 (R Core Team, 2023); for plotting, the packages *ggplot2* version 3.4.4 (Wickham, 2016), *ggpmisc* version 0.5.5 (Aphalo, 2022), and *questionr* version 0.7.8 to identify all R packages used.

2.3.1 | Proportion of papers sharing code over time

We tested for a trend in code-sharing over time by modeling code sharing (binary, yes/no) as a function of the year (relative to 2010) using a generalized linear model. Modeling was performed using the function *glm* in the *stats* R package (R Core Team, 2023) with a binomial error distribution. We similarly tested for temporal trends in two other open-science components, open-access publication (binary) and open data (binary). We also tested whether open-access or open-data papers were disproportionately likely to share code using chi-square tests via the *chisq.test* function in the *stats* R package (R Core Team, 2023).

2.3.2 | Impact of code sharing on citations

We additionally modeled the relationship between code sharing and citation count using generalized linear models in R. We modeled the dependent variable (cumulative number of citations of each article by 2022) using a Poisson distribution, which models the number of independent events occurring within a period of time (Bolker, 2008). In addition to the predictor variable for code sharing (binary, yes/no), we included other variables that were hypothesized to influence citation count. Data sharing (binary, yes/no) may increase citation counts as readers may cite papers as data sources (Christensen et al., 2019; Piwowar et al., 2007). Open access (binary, yes/no) may also increase citation counts by reaching a broader set of readers (Tang et al., 2017). Publications accumulate citations over time, and

so citation count should increase with publication age (continuous, 1–13 years). Finally, publications in higher impact journals may be more likely to be read and cited, and hence, journal impact factor (continuous, 0–11.633) may be positively associated with citation count. In addition to main effects, we considered two classes of interactions: (1) interactions between publication age and other main effects, which are appropriate if a main effect modifies the rate at which a publication accumulates citations over time; and (2) interactions between open-science criteria (i.e., open access, open code, and open data), which are appropriate if there are synergistic effects of meeting multiple open-access criteria. We compared 11 models (including one null model) that differed in complexity and that represented different hypotheses regarding the factors that influence citations (Table 1). Continuous variables were scaled and centered. Overall model pseudo- R^2 for the best-performing model was calculated using the function *r.squaredGLMM* in the *rsq* package (Zhang, 2018).

3 | RESULTS

We identified 28,227 articles that met our search criteria. From this set of articles, we randomly selected 1001 papers (the closest number to 1000 that is evenly divisible by 13) evenly spread across the temporal range (13 years) for a total of 77 papers per year. Overall, R code was only available for 55 of the 1001 papers examined (5.5%; Figure 1). When shared, code was most often in the Supplemental Information (40%), followed by Github (22%), Figshare (11%), or other repositories (37%). The majority of code (67%) did not include a license. Where a license was included, it was nearly always permissive or copyleft (e.g., CC0, CC-BY, GPL, and MIT), with only one publication including a proprietary license. Open-access publications were twice as likely to share code than closed-access publications (8.5% vs 4.24%, $\chi^2=7.2576$, $p=.008599$). Publications with open data were 12 times more likely to share code than closed-data publications (26.5% vs 2.2%, $\chi^2=133.36$, $p=9.999e-05$). Among the set of publications that did not cite R, 6.2% mentioned using R in the text. Of the 33 papers that shared code via potentially citable DOIs or URLs, we were unable to find any citations of the code itself.

3.1 | Code sharing over time

The proportion of publications sharing code has increased significantly ($p=.00157$) over time (Figure 1, Table 2), with code sharing increasing at an average of 0.6% per year over this period. A Durbin-Watson test indicated no temporal autocorrelation in residuals ($DW=1.7544$, $p=.6475$). We note that the years of 2021 and 2022 showed notable shifts toward more frequent sharing (although 2013 showed a similar level of code sharing), but the percentage of code sharing has been consistently below 20% over the past decade, and has remained lower than the percentage of open-access papers or papers sharing data (Figure 1). Over this same time period, the

TABLE 1 Candidate models of citation count.

ID	Models	df	Δ AIC
1	Citations ~ Impact factor \times Age + R code shared \times Age + Open access \times Age + Data shared \times Age + Data shared \times R code shared + R code shared \times Open Access + Open access \times Data shared	13	0.0
2	Citations ~ Impact factor \times Age + R code shared \times Age + Data shared \times Age + Data shared \times R code shared	9	834.0
3	Citations ~ Impact factor \times Age + R code shared \times Age + Open access \times Age + Open access \times R code shared	9	907.5
4	Citations ~ Impact factor \times Age + R code shared \times Age + Open access \times R code shared	8	913.0
5	Citations ~ Impact factor \times Age + R code shared \times Age + Open access \times Age + Data shared \times Age	10	937.8
6	Citations ~ R code shared + Data shared + Open access + Age + Impact factor	6	1116.9
7	Citations ~ Impact factor \times Age + R code shared \times Age + Open access \times Age	8	1181.5
8	Citations ~ Impact factor \times Age + R code shared \times Age	6	1339.7
9	Citations ~ Impact factor \times Age	4	1973.8
10	Citations ~ Age	2	5824.3
11	Citations ~1	1	13721.6

proportion of publications including data also increased significantly ($p = 1.48e-06$; Figure 1), while the proportion of open-access publications did not change significantly ($p = .926$; Figure 1).

3.2 | Impact of code sharing on citations

Our best-performing model of citation counts (ID 1; Table 1) had a substantially better AIC score than our other candidate models (Δ AIC ≥ 834.0). Impact factor ($\beta = .326398$, $p < 2e-16$), age ($\beta = .626261$, $p < 2e-16$), and open access ($\beta = .069522$, $p = .000628$), all had significantly positive associations with the number of

citations, while code availability ($\beta = -1.425857$, $p < 2e-16$) and open data ($\beta = -.07596$, $p = .025014$) both showed significantly negative associations (Table 3). Interactions between age and both code availability ($\beta = .562521$, $p < 2e-16$) and data availability ($\beta = .212861$, $p < 2e-16$) were positive and significant, while interactions between age and both impact factor ($\beta = .001585$, $p = .803079$) and open access ($\beta = .022175$, $p = .20503$) were not significant. Pairwise interactions between the three open-access criteria were all positive and significant: code availability and data availability ($\beta = 1.468237$, $p < 2e-16$), code availability and open access ($\beta = 1.454602$, $p < 2e-16$), and open access and data availability ($\beta = .174743$, $p = 5.08e-05$). The overall model had a pseudo- R^2 of .93.

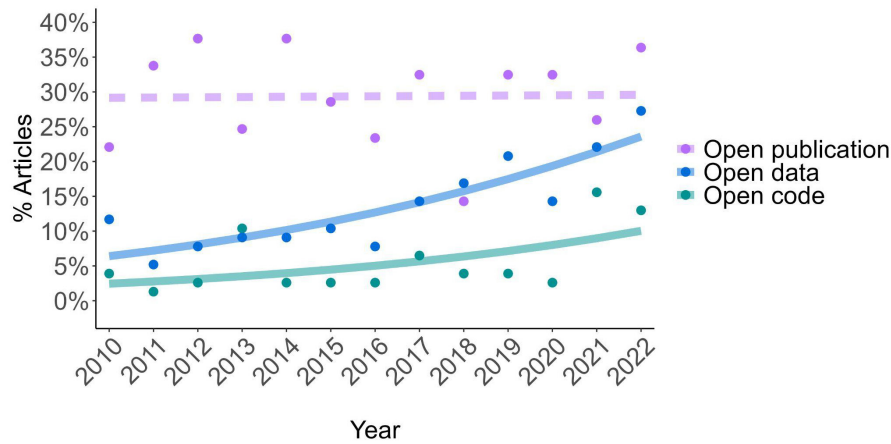


FIGURE 1 Temporal trends of open-access publications (open publication), open data, and code sharing (open code) between 2010 and 2022. Sample size was 77 papers per year, 1001 total. The lines show trends in each open-science component over time, with dashed lines representing non-significant trends, and solid lines indicating significance. Open data ($p = 1.48e-06$) and code sharing ($p = .00157$) increased significantly over time, while open-access publications ($p = .926$) had no statistically significant relationship. Note that publications may be converted to open access after publication as journal policies change, so the open-publication trend should be interpreted with caution.

TABLE 2 Estimated coefficients for models of temporal trends in code sharing, data sharing, and open access over time.

Responses	Parameters	Estimates	Std. error	z-Value	p-Value
Code sharing	Intercept	-3.68612	0.32626	-11.298	<2e-16***
	Year (relative to 2010)	0.12443	0.03936	3.161	.00157**
Data sharing	Intercept	-2.68284	0.21073	-12.731	<2e-16***
	Year (relative to 2010)	0.12561	0.02609	4.814	1.48e-06***
Open access	Intercept	-0.88778	0.13130	-6.761	1.37e-11***
	Year (relative to 2010)	0.00172	0.01855	0.093	.926

*** $p < .001$, ** $p < .01$.

4 | DISCUSSION

We found that the scientific literature in ecology and evolution still falls far short of the code sharing required for adequate reproducibility and transparency, despite an increasing trend in code sharing over the last 12 years. This low rate of code sharing undoubtedly hinders scientific progress and likely has far-reaching financial consequences, since a lack of reproducibility means that coding must be continuously redone for common analytical tasks (Freedman et al., 2015). Further, our results indicate that the failure to share code may also reduce the academic impact of scientists, as sharing code leads to a higher rate of citation accumulation (i.e., a significant year-by-code-sharing interaction; $p < 2e-16$; Table 3). The positive effect of code sharing is increased when other open-science criteria are also met, as evidenced by significant, positive interactions between open access and code sharing ($p < 2e-16$; Table 3) and between data sharing and code sharing ($p < 2e-16$; Table 3), with publications meeting all three open science criteria (code sharing, data sharing, and open access) having the highest overall predicted citation rates ("Fully open," Figure 2). Surprisingly, our model found negative effects of code sharing and data sharing on citation

count, despite the positive impacts of code sharing and data sharing on citation rate (i.e., significant interactions between code or data sharing and age). One possible cause of this discrepancy could be that scientists may be less likely to share code and data underlying publications they expect to be impactful if they are planning related studies using the same code or data. Alternatively, this discrepancy may be due to the increase in code availability with time leading to a disproportionate amount of the papers which share code being young, and hence having few citations. We also did not find support for an interaction between publication age and impact factor, suggesting that though impact factor may affect the total number of citations, it does not strongly affect the citation rate. We also did not find evidence of a significant interactive effect of publication age and open access on citation count. This failure of open access to impact citation rate, as well as the lack of change in the percentage of open access publications over time (Figure 1) may be driven by changes to open access status of publications over the years. As journals switch to open access, many papers that were initially published closed access have been converted to open after varying lengths of time, potentially degrading the signal of open access on citation counts.

	Estimate	Std. error	Z-value	Pr(> z)
(Intercept)	2.696757	0.011259	239.52	<2e-16***
Impact factor	0.326398	0.006472	50.429	<2e-16***
Age	0.626261	0.010294	60.837	<2e-16***
R code shared (yes)	-1.425857	0.094131	-15.148	<2e-16***
Open access (yes)	0.069522	0.020333	3.419	.000628***
Data shared (yes)	-0.07596	0.033893	-2.241	.025014*
Impact factor: Age	0.001585	0.006358	0.249	.803079
Age: R code shared (yes)	0.562521	0.037058	15.179	<2e-16***
Age: Open access (yes)	0.022175	0.017498	1.267	.20503
Age: Data shared (yes)	0.212861	0.02383	8.932	<2e-16***
R code shared (yes): Data shared (yes)	1.468237	0.069624	21.088	<2e-16***
R code shared (yes): Open access (yes)	1.454602	0.070556	20.616	<2e-16***
Open access (yes): Data available (yes)	0.174743	0.043125	4.052	5.08e-05***

Note: Statistically significant model terms in bold. *** $p < .001$, * $p < .05$.

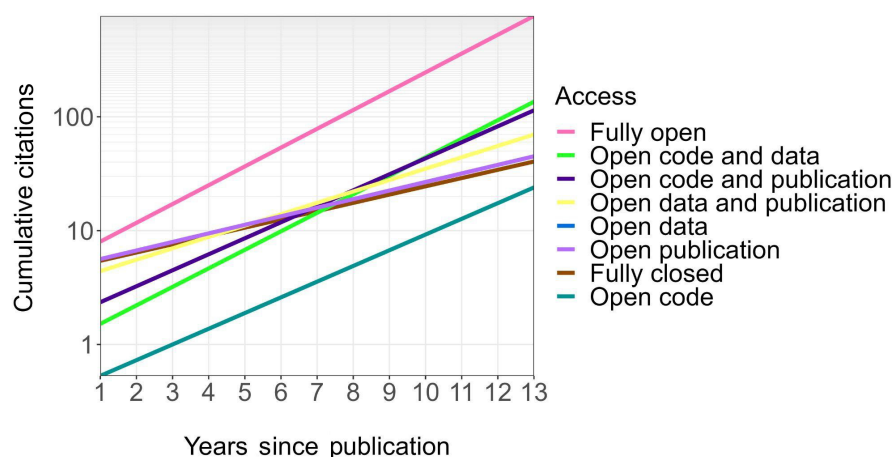


TABLE 3 Estimated coefficients for the selected model predicting number of publication citations (ID 1 in Table 1). Bold entries indicate significance at $\alpha = .05$.

FIGURE 2 Predicted impacts of code sharing on cumulative citations. Predicted values shown are for the mean impact factor (3.0) across the publications analyzed. Fully open = open code, open data, and open-access publication; fully closed = closed-access publication, full data not shared, and no publicly available code. Predictions are based on estimated model coefficients (Table 3). Note that the open-data line is hidden by the open-publication line.

Our results suggest that scientists and journals who embrace code sharing may be more impactful, even if the publications are closed-access. For instance, our model predicts that a scientist who publishes a paper sharing code and data in a journal with an impact factor of 3 will have approximately double the citations after 11 years (64.1 vs 29.0) than if they had published without sharing code and data (Figure 2). However, if this same paper is also published open-access, our model predicts it will have twice as many citations after only 3 years (17.1 vs 7.6, Figure 2). Further, although papers in high-impact journals may be expected to receive more citations in general than those in low-impact journals (although note that the impact factor by age interaction was not significant), our model suggests code sharing may balance this out, particularly when combined with other open-science factors. By the third year following publication, a fully open publication in a low-impact (10th percentile, impact factor = 1.3) journal was predicted to have more citations than a fully closed publication in a high-impact (90th percentile, impact factor = 4.7) journal (12.3 vs 10.6, Figure 3). For publications in low-impact journals sharing code and meeting only one

other open science criterion (open access or data sharing), the fully closed publication in a high-impact journal is predicted to be surpassed by the 11th year (45.8 and 42.6 vs 40.5; Figure 3).

Our finding of a significant effect of code sharing on citation rate is surprising given that most code was not licensed for reuse. Software licenses play a critical (and often unappreciated) role in code sharing: permissive licenses (e.g., MIT) encourage reuse, while restrictive or proprietary licenses can still allow methodological transparency while limiting or preventing the reuse of published code (Stodden, 2009). Importantly, where code is published without a license, the author retains the copyright (Stodden, 2009). The finding of a significant effect of code sharing on citation rate despite the rarity of permissive software licenses may indicate that the citation benefits for code-sharing publications are due to either (1) code being reused without proper licensing, or (2) availability of code increasing citations in the absence of code reuse. The latter could happen if code availability increased confidence in the findings of a publication. Given the increasing importance of code in ecology and evolution (Feng et al., 2020), both scientists and funding

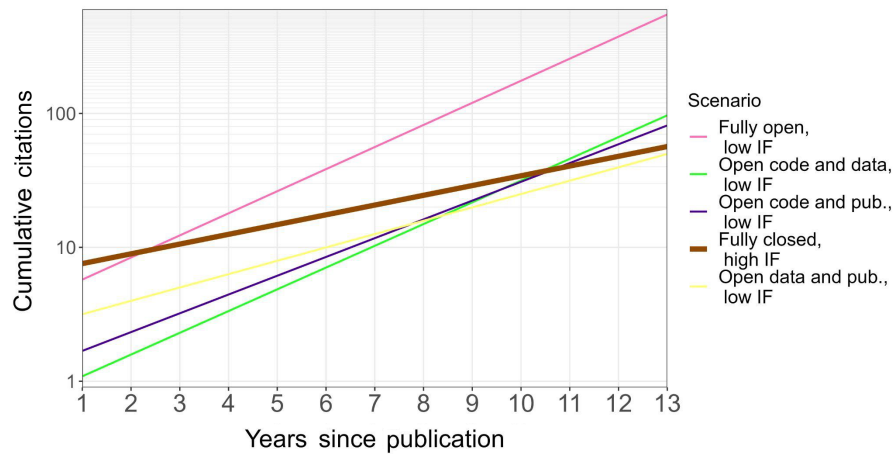


FIGURE 3 Code sharing can balance out low impact factors. Low impact factor (IF; 1.3) and high impact factor (4.7) are defined using the 0.1 and 0.9 quantiles in our dataset. Fully open = open code and open access publication; fully closed = closed-access publication and lack of publicly available code. Predictions are based on estimated model coefficients (Table 3). Our model predicts that 3 years after publication, fully open papers published in a low impact journal may have roughly the same number of citations as fully closed papers published in high impact journals (on average). Legend is arranged in descending order of citations at year 13.

organizations need to give more thought to software licenses. We note that considering licensing is important both for scientists who wish their code to be freely available and benefit from citations stemming from reuse as well as for scientists who wish to embrace transparency without allowing use of their code.

We note that there are important limitations to our study. The low rates of code sharing and data sharing limited our sample sizes which in turn likely impacted model precision. These low sample sizes may help explain the anomalously high number of publications sharing code in 2013 (Figure 1). Further, these low rates of sharing led to a major imbalance in our main variable of interest, code sharing (55 that shared vs 946 that did not). We also note that while we treated code and data sharing as binary variables, there is a tremendous amount of variation in the amount and quality of data and code that are shared. Where some publications include only summary data or example code, others include well-documented code and data, and this variation could impact citation count. Some of the variables we examined may change over time, potentially weakening inferences: archived data and code may be lost, publications may become open access, and impact factors change over time. Our search was limited to English language publications, so these trends may not hold for publications in other languages (Konno et al., 2020), though they may be expected to share code even less often (Serwadda et al., 2018). Further, our work focused on papers that cited R, and does not account for the small percentage of papers which note using R but do not cite it. Thus, our results may overestimate the proportion of total publications around the globe sharing code. Finally, our work focused on one programming language, R, which may not be broadly representative.

To create an environment conducive to reproducibility and transparency, we call upon scientists, funding sponsors, publishers, and institutions to champion code sharing and acknowledge its role as a valuable and necessary contribution to the scientific process. Major funding sponsors are beginning to mandate open science approaches

(e.g., the EU's Horizon Europe: https://research-and-innovation.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en, and the US Year of open science: <https://open.science.gov/>). Scientific journals can also play a significant role in this transformation (McNutt, 2014; Mislan et al., 2016; Nosek et al., 2015; "Reality check on reproducibility", 2016). Minimally, journals facilitate the deposition of code in a stable repository and provide a link to that repository (Mislán et al., 2016; ideally in both human and machine-readable formats; Peng, 2011). Unfortunately, even when funders and journals mandate data and code sharing, compliance often remains low (Culina et al., 2020). More ambitious solutions might include incorporating links between methods text and the corresponding code, employing dedicated code editors to help improve code style and clarity (similar to Data Editors employed by The American Naturalist), or incorporating computational notebooks (e.g., RMarkdown, Quarto, Jupyter; Peng, 2011). Such measures will enhance transparency in reporting and provide reviewers and readers with the critical information necessary to reproduce and validate the study's findings. The adoption of these principles and practices will serve to promote the integration of open code in the scientific landscape, enhancing the verifiability and impact of our research.

Given the growing list of reasons for code sharing, we encourage scientists to embrace open code and open science more generally. Although maintaining well-documented code in a version-controlled public repository (e.g., Github) and public archive with directions for its use (e.g., Zenodo) is ideal for code sharing, other options that require less effort can at least ensure the distribution of code to other researchers interested in using it. Recent advances in artificial intelligence (e.g., ChatGPT) have made documenting scripts easier, thus lowering the cost to authors to share documented code (Merow, Serra-Diaz, et al., 2023). Finally, we stress that as code sharing increases, our attribution practices must keep pace, both for scientific transparency and to credit the developers (Merow, Boyle, et al., 2023).

AUTHOR CONTRIBUTIONS

Brian Maitner: Conceptualization (lead); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); project administration (lead); resources (lead); software (lead); supervision (equal); visualization (lead); writing – original draft (lead); writing – review and editing (equal). **Paul Efren Santos Andrade:** Investigation (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Luna Lei:** Formal analysis (supporting); investigation (supporting); methodology (supporting); software (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Jamie Kass:** Conceptualization (supporting); methodology (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Hannah L. Owens:** Conceptualization (supporting); methodology (supporting); writing – original draft (supporting); writing – review and editing (supporting). **George C. G. Barbosa:** Writing – original draft (supporting); writing – review and editing (supporting). **Brad Boyle:** Conceptualization (supporting); methodology (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Matiss Castorena:** Conceptualization (supporting); methodology (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Brian J. Enquist:** Conceptualization (supporting); methodology (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Xiao Feng:** Conceptualization (supporting); methodology (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Daniel S. Park:** Conceptualization (supporting); methodology (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Andrea Paz:** Conceptualization (supporting); methodology (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Gonzalo Pinilla-Buitrago:** Conceptualization (supporting); methodology (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Cory Merow:** Conceptualization (supporting); methodology (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Adam Wilson:** Conceptualization (supporting); funding acquisition (lead); methodology (supporting); project administration (lead); supervision (lead); writing – original draft (supporting); writing – review and editing (supporting).

AFFILIATIONS

- ¹Department of Integrative Biology, University of South Florida, St. Petersburg, Florida, USA
- ²Department of Geography, University at Buffalo, Buffalo, New York, USA
- ³Departamento de Ecología, Universidad Nacional San Antonio, Abad del Cusco, Cusco, Peru
- ⁴University at Buffalo, Buffalo, New York, USA
- ⁵Macroecology Laboratory, Graduate School of Life Sciences, Tohoku University, Sendai, Miyagi, Japan
- ⁶Center for Global Mountain Biodiversity, Globe Institute, University of Copenhagen, Copenhagen, Denmark
- ⁷Center for Macroecology, Evolution, and Climate, Globe Institute, University of Copenhagen, Copenhagen, Denmark
- ⁸Florida Museum of Natural History, University of Florida, Gainesville, Florida, USA
- ⁹Department of Computer Science, University of Arizona, Tucson, Arizona, USA
- ¹⁰Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona, USA

- ¹¹The Santa Fe Institute, Santa Fe, New Mexico, USA
- ¹²Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
- ¹³Department of Biological Sciences, Purdue University, West Lafayette, Indiana, USA
- ¹⁴Purdue Center for Plant Biology, Purdue University, West Lafayette, Indiana, USA
- ¹⁵Department of Environmental Systems Science, Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland
- ¹⁶Département de Sciences Biologiques, Université de Montréal, Montreal, Québec, Canada
- ¹⁷PhD Program in Biology, The Graduate Center and Department of Biology, City College of New York, City University of New York, New York, New York, USA
- ¹⁸Eversource Energy Center and Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut, USA

ACKNOWLEDGMENTS

H.L.O. was supported under EU Horizon 2020 Marie Skłodowska-Curie grant agreement No. 891702 and VILLUM FONDEN research grant No. 25925. A.M.W. and B.M. were supported by NASA BioSCape (80NSSC21K0086) and NASA Ecological Forecasting Team Applied Sciences Program (80NSSC21K1183). B.J.E. and C.M. were supported by NSF awards Nos 2225076 and 2225078.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

OPEN RESEARCH BADGES



This article has earned Open Data and Open Materials badges. Data and materials are available at <https://doi.org/10.5281/zenodo.8201251>.

DATA AVAILABILITY STATEMENT

All code and data underlying this work are available on Github at https://github.com/bmaitner/R_citations and are permanently archived at <https://doi.org/10.5281/zenodo.8201251>.

ORCID

Brian Maitner <https://orcid.org/0000-0002-2118-9880>

Paul Efren Santos Andrade <https://orcid.org/0000-0002-6635-0375>

Jamie Kass <https://orcid.org/0000-0002-9432-895X>

Andrea Paz <https://orcid.org/0000-0001-6484-1210>

REFERENCES

- Aphalo, P. J. (2022). *ggpmisc: Miscellaneous Extensions to "ggplot2"*. <https://CRAN.R-project.org/package=ggpmisc>
- Baker, M. (2016). 1500 scientists lift the lid on reproducibility. *Nature*, 533, 452–454.
- Barton, C. M., Lee, A., Janssen, M. A., van der Leeuw, S., Tucker, G. E., Porter, C., Greenberg, J., Swantek, L., Frank, K., Chen, M., & Jagers, H. R. A. (2022). How to make models more useful. *Proceedings of the National Academy of Sciences of the United States of America*, 119, e2202112119.
- Barton, K. (2023). *MuMIn: Multi-model inference*. <https://CRAN.R-project.org/package=MuMIn>

- Bolker, B., & R Development Core Team. (2022). *bbmle: Tools for general maximum likelihood estimation*. <https://CRAN.R-project.org/package=bbmle>
- Bolker, B. M. (2008). *Ecological models and data in R*. Princeton University Press.
- Bryan, J. (2023). *googlesheets4: Access Google Sheets using the Sheets API V4*.
- Busjahn, T., & Schulte, C. (2013). The use of code reading in teaching programming. In *Proceedings of the 13th Koli calling international conference on computing education research, Koli Calling 13* (pp. 3–11). Association for Computing Machinery.
- Cadwallader, L., & Hrynaskiewicz, I. (2022). A survey of researchers' code sharing and code reuse practices, and assessment of interactive notebook prototypes. *PeerJ*, 10, e13933.
- Christensen, G., Dafoe, A., Miguel, E., Moore, D. A., & Rose, A. K. (2019). A study of the impact of data sharing on article citations using journal policies as a natural experiment. *PLoS One*, 14, e0225883.
- Culina, A., van den Berg, I., Evans, S., & Sánchez-Tójar, A. (2020). Low availability of code in ecology: A call for urgent action. *PLoS Biology*, 18, e3000763.
- D'Agostino McGowan, L., & Bryan, J. (2023). *Googledrive: An interface to google drive*. <https://CRAN.R-project.org/package=googledrive>
- Drescher, M., & Edwards, R. C. (2019). A systematic review of transparency in the methods of expert knowledge use. *Journal of Applied Ecology*, 56, 436–449.
- Feng, X., Qiao, H., & Enquist, B. J. (2020). Doubling demands in programming skills call for ecoinformatics education. *Frontiers in Ecology and the Environment*, 18, 123–124.
- Freedman, L. P., Cockburn, I. M., & Simcoe, T. S. (2015). The economics of reproducibility in preclinical research. *PLoS Biology*, 13, e1002165.
- Gomes, D. G. E., Pottier, P., Crystal-Ornelas, R., Hudgins, E. J., Foroughirad, V., Sánchez-Reyes, L. L., Turba, R., Martínez, P. A., Moreau, D., Bertram, M. G., Smout, C. A., & Gaynor, K. M. (2022). Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings of the Biological Sciences*, 289, 20221113.
- Hartig, F. (2022). *DHARMA: Residual diagnostics for hierarchical (multi-level/mixed) regression models*. <https://CRAN.R-project.org/package=DHARMA>
- Keirstead, J. (2016). *Scholar: Analyse citation data from Google Scholar*. <https://github.com/jkeirstead/scholar>
- Konno, K., Akasaka, M., Koshida, C., Katayama, N., Osada, N., Spake, R., & Amano, T. (2020). Ignoring non-English-language studies may bias ecological meta-analyses. *Ecology and Evolution*, 10, 6373–6384.
- Lai, J., Lortie, C. J., Muenchen, R. A., Yang, J., & Ma, K. (2019). Evaluating the popularity of R in ecology. *Ecosphere*, 10, e02567.
- Maitner, B. S., & Lei, L. (2024). *bmaitner/R_citations: Revised Pre-print release*. <https://zenodo.org/doi/10.5281/zenodo.8201250>
- Mauranen, A., Hynninen, N., & Ranta, E. (2010). English as an academic lingua franca: The ELFA project. *English for Specific Purposes*, 29, 183–190.
- McNutt, M. (2014). Journals unite for reproducibility. *Science*, 346, 679.
- Merow, C., Boyle, B., Enquist, B. J., Feng, X., Kass, J. M., Maitner, B. S., McGill, B., Owens, H., Park, D. S., Paz, A., Pinilla-Buitrago, G. E., Urban, M. C., Varela, S., & Wilson, A. M. (2023). Better incentives are needed to reward academic software development. *Nature Ecology & Evolution*, 7, 626–627.
- Merow, C., Serra-Diaz, J. M., Enquist, B. J., & Wilson, A. M. (2023). AI chatbots can boost scientific coding. *Nature Ecology & Evolution*, 7, 960–962.
- Mislán, K. A. S., Heer, J. M., & White, E. P. (2016). Elevating the status of code in ecology. *Trends in Ecology & Evolution*, 31, 4–7.
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1, 0021.
- Muschelli, J. (2019). *rscopus: Scopus database "API" interface*. <https://CRAN.R-project.org/package=rscopus>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348, 1422–1425.
- Parker, T. H., Forstmeier, W., Koricheva, J., Fidler, F., Hadfield, J. D., Chee, Y. E., Kelly, C. D., Gurevitch, J., & Nakagawa, S. (2016). Transparency in ecology and evolution: Real problems, real solutions. *Trends in Ecology & Evolution*, 31, 711–719.
- Parker, T. H., Nakagawa, S., Gurevitch, J., & IIEE (Improving Inference in Evolutionary Biology and Ecology) workshop participants. (2016). Promoting transparency in evolutionary biology and ecology. *Ecology Letters*, 19, 726–728.
- Peng, R. D. (2011). Reproducible research in computational science. *Science*, 334, 1226–1227.
- Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS One*, 2, e308.
- Public Library of Science. (2023). *PLOS open science indicators*. <http://dx.doi.org/10.6084/M9.FIGSHARE.21687686.V3>
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Reality Check on Reproducibility. (2016). Reality check on reproducibility. *Nature*, 533, 437.
- Serwadda, D., Ndebele, P., Grabowski, M. K., Bajunirwe, F., & Wanyenze, R. K. (2018). Open data sharing and the global south—who benefits? *Science*, 359, 642–643.
- Stodden, V. (2009). The legal framework for reproducible scientific research: Licensing and copyright. *Computing in Science & Engineering*, 11, 35–40.
- Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 2584–2589.
- Tang, M., Bever, J. D., & Yu, F.-H. (2017). Open access increases citations of papers in ecology. *Ecosphere*, 8, e01887.
- van der Loo, M. P. J. (2014). The stringdist package for approximate string matching. *The R Journal*, 6, 111–122.
- Wagenmakers, E.-J., Sarafoglou, A., Aarts, S., Albers, C., Algermissen, J., Bahnik, Š., van Dongen, N., Hoekstra, R., Moreau, D., van Ravenzwaaij, D., Sluga, A., Stanke, F., Tendeiro, J., & Aczel, B. (2021). Seven steps toward more transparency in statistical practice. *Nature Human Behaviour*, 5, 1473–1480.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. <https://ggplot2.tidyverse.org>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4, 1686.
- Zhang, D. (2023). *rsq: R-squared and related measures*. R package version 2.6. <https://CRAN.R-project.org/package=rsq>

How to cite this article: Maitner, B., Santos Andrade, P. E., Lei, L., Kass, J., Owens, H. L., Barbosa, G. C. G., Boyle, B., Castorena, M., Enquist, B. J., Feng, X., Park, D. S., Paz, A., Pinilla-Buitrago, G., Merow, C., & Wilson, A. (2024). Code sharing in ecology and evolution increases citation rates but remains uncommon. *Ecology and Evolution*, 14, e70030. <https://doi.org/10.1002/ece3.70030>