Revision round #1

Decision for round #1 : *Revision needed*

I welcome this formal analysis of reproducibility practices in Ecology. In general, I don't have any major concerns with the paper, but both reviewers raise important points that need to be addressed before recommending the manuscript. Some of those comments refer to issues that might not be easily fixed, such as the relatively smaller sample size or the fact that the study does not cover more recent time periods. Still, a good discussion should be included regarding those points. However, other points, such as the revision of the supplementary material of the analyzed manuscripts, are appropriate and will reinforce the message of the paper.

I am looking forward to reading an enhanced version of the paper.

Best,

Ignasi Bartomeus.

Response 1: We thank the recommender for the positive comments on our preprint and the suggestions for how to deal with some of the criticism of the reviewers. We have now implemented the requested changes and provided a much-improved version of this preprint, which we hope will satisfy the reviewers and the recommender. The new version of the preprint has been updated and is now available (please note that during the re-submission, we found a couple of minor typos which we will fix for the next preprint update; for the time being, we have attached to our response the version with those typos corrected. Sincere apologies).

Review by Francisco Rodriguez-Sanchez, 21 Jan 2025 13:07

Title and abstract

Does the title clearly reflect the content of the article? YES

Does the abstract present the main findings of the study? YES

Introduction

Are the research questions/hypotheses/predictions clearly presented? YES

Does the introduction build on relevant research in the field? YES

Materials and methods

Are the methods and analyses sufficiently detailed to allow replication by other researchers? YES (although the list of examined papers is not provided)

Response 2: We did not find the comment regarding the missing list of examined papers below, so we are unsure what exact list the reviewer refers to. We believe we have provided all the datasets needed in the folder "data" (https://github.com/ASanchez-Tojar/code-sharing_policies_matter/tree/main/data). We think that the reviewer refers to the randomly selected papers (2015-2016: 200 references, 2018-2019: 200), which are provided here: https://github.com/ASanchez-Tojar/code-sharing_policies_matter/tree/main/data/raw_databases/no_policies_screening, but please do let us know otherwise.

Are the methods and statistical analyses appropriate and well described? YES, but see comments below

Results

Are the results described and interpreted correctly? YES, but see comments below

Discussion

Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? YES

Are the conclusions adequately supported by the results (without overstating the implications of the findings)? YES

Sánchez-Tójar et al. present an interesting analysis of how journal policies on code sharing may affect levels and quality of actual code sharing, determining the degree of computational reproducibility of published scientific results in ecology. This is an important topic and a nice follow-up from Culina et al. (2020) which analysed the computational reproducibility in ecological journals encouraging or mandating code sharing. Here, those results are contrasted against a comparable sample of ecological journals without a code sharing policy, finding that, in fact, journals' code sharing policies seem to increase actual code sharing and computational reproducibility.

Response 3: We thank the reviewer for the positive comments and great constructive criticism, which has helped us improve our preprint.

The comparison with Culina et al. (2020) implies that papers from the same years (2015 to 2019) were analysed here. More than five years have passed now, which is a long time in the open science world, in the sense that actual code sharing practices may have

changed now, and be significantly different from the figures reported here (from a few years ago). This is fine and justified by the planned comparison with Culina et al. (2020), with the main goal of assessing the impact of journal policies on code sharing. Other pieces of work (e.g. Maitner et al. 2024 https://doi.org/10.1002/ece3.70030, and others to come) may provide more updated assessments of code sharing rates and quality.

Response 4: We agree about the time passed. We planned to have this preprint ready much earlier, but we unfortunately could not. We carefully considered whether we should update our searches but decided that for the sake of full comparison with Culina et al. (2020)'s dataset, we should avoid it. We have now tried to be more explicit to avoid readers misinterpreting our estimates as current ones and lead readers to the most recent surveys on the topic, and have added the following to the first paragraph of the methods.

"*In addition, readers should be aware that our study was designed to allow us to fully compare in detail our results for journals without a code-sharing policy to those with from Culina et al. (2020), so we restricted our searches from 2015 to 2019. Subsequently, the general figures on data- or code-sharing and reporting practices reported here might not be representative of current ones, for which we refer to recent surveys (Maitner et al. 2024; Sperandii et al. 2024).*"

Main comments

(1) L161-162: "we searched for software and package versions not only within the text but also in the reference list of the corresponding article".

I think this methodology may be underestimating the numbers on the reporting of software and package versions, as I think many articles report that information in the Supplementary Material (e.g. Extended Methods section, or Software Appendix), or even within the archived code, rather than within the main text or reference list. Hence, I think it may be important to revise at least the Supplementary Information of all the examined papers (where existing) to check if software and packages are reported there. To be fully comprehensive, provided code should be revised as well, for the few papers that provide it. Numbers might not change much compared to current figures, but I am afraid we will not know unless we check.

Response 5: We have clarified this further as we did indeed revise the supplementary text whenever available (not so the scripts or readme's though). Here is the changed we have consequently implemented in the text:

"*we searched for software and package versions within the main article and supplementary text (whenever available), including in their reference lists*"

Indeed, we had not searched for the software and package versions in the code, and we prefer not to do so because (1) we think it is important to focus on what is reported in the main text and its supplements so that numbers are comparable across articles that report

none, some or all code, particularly since we agree with the suggestion from the reviewer regarding that only fundamental software and packages should be reported in the main manuscript and its supplements (see *Response 6*); (2) even if we were to explore the code, we would still underestimate the number of packages used unless we ran the code since some packages might not be attached or required within the scripts. We think that checking all the corresponding scripts would not be a cost-effective approach for our purposes. For the sake of transparency, here are the numbers regarding articles that have reported at least some versions of those that have shared at least some code:

- With policy: 47 out of the 89 articles reported the version of some or all software, and 32 out of 65 reported some or all packages
- Without policy: 8 out of the 15 articles reported the version of some or all software, and 6 out of 13 reported some or all packages

(2) L290-303: "We found that versions of the statistical software and packages were often missing, and about a tenth of the articles did not even state the software used. Reporting software and package versions is important for several reasons. First, they can help in understanding and solving technical issues related to software dependencies, which are one of the most often encountered factors hindering computational reproducibility (Laurinavichyute et al. 2022; Kellner et al. 2024; Samuel and Mietchen 2024). Different versions of software and/or packages can lead to inconsistencies in results and even to code rot, which occurs when code relies on specific versions of software or packages that are no longer available or have undergone significant changes (e.g., deprecated functions), rendering the code incompatible with current operating systems (Boettiger 2015; Laurinavichyute et al. 2022)."

This is related to the comment above. The authors seem to expect (and recommend) that software and package versions are reported in the main text and reference list to guarantee computational reproducibility. But is it feasible or reasonable to expect that all software dependencies are cited in the main text, to guarantee computational reproducibility? A typical analysis in ecology may involve dozens, or even more than a hundred, software dependencies (including indirect dependencies).

Response 6: This is a great point, and we agree we cannot and should not expect that all packages (including dependencies) are included in the main manuscript or the supplementary material (see also *Response 7*). Indeed, assessing whether all software and packages used (or even focusing only on the key analytical packages) are reported would already be a large enterprise in itself. That is why we focused on simply compiling information on how many analytical software and packages were reported, and from those, calculating the percentage of software and packages for which the version was provided. We think those estimates provide a good understanding of software/package reporting, particularly since few studies shared any code. In addition, we thought it would be quite informative to share the number of studies that did not report any statistical software despite reporting statistical analyses to showcase an extreme case of low reporting.

Current guidelines for software citation (Chue Hong et al. 2019, https://doi.org/10.5281/zenodo.3479198) state that "you should cite software that has a significant impact on the research outcome presented in your work, or on the way the research has been conducted. In general, you do not need to cite software packages or libraries that are not fundamental to your work and that are a normal part of the computational and scientific environment used. These dependencies do not need to be cited outright but should be documented as part of the computational workflow for complete reproducibility".

Response 7: We agree with this suggestion, and have now modified the following text in the introduction to make this point clearer, as well as to clarify why we focused on reported software and packages.

"*At the minimum, the software and packages that are fundamental for the analyses should be stated and appropriately referenced in the main manuscript, and the version(s) used should be clearly stated (guidelines for software citation: Chue Hong et al. 2019). The remaining packages should be documented as part of stand-alone documentation (e.g., README, or inline comments; Benureau and Rougier 2018; Jenkins et al. 2023; Ivimey-Cook et al. 2023). Software and package citation is important for computational reproducibility, but also to better explain the methodology and give credit to software developers.*"

At the minimum, the software and packages used that are fundamental for the analyses should be stated and appropriately referenced in the main manuscript, and the version(s) used should be clearly stated (guidelines for software citation: Chue Hong et al. 2019). The remaining packages should be documentedin the manuscript and/or as part of stand-alone documentation (e.g., README, or inline comments; Benureau and Rougier 2018; Jenkins et al. 2023; Ivimey-Cook et al. 2023). Software and package citation is not only important for computational reproducibility, but also to better explain the methodology and give credit software developers.

Thus, the main function of software citations in the reference list is not to guarantee computational reproducibility, but to explain methods and give credit to software developers.

Response 8: Absolutely! We have added that clarification to the introduction (see response 7) as well as the following text to the discussion:

"*In addition, we advocate for proper software and package citation to give credit to software developers and incentivise software and package development (e.g., using the R package 'grateful'; Rodriguez-Sanchez and Jackson 2024).*"

For the sake of computational reproducibility, a better way of reporting dependencies is including a full report of the computational environment used for the analyses, including, of course, package versions. It would be nice, in my opinion, if this manuscript included brief guidelines on how to do so within the final recommendations, or at least point to

useful tutorials. For example, focusing on R, authors could include an appendix generated with `sessionInfo()` or `grateful::cite_packages()`, or include in their code a file reporting the computational environment (e.g. a Dockerfile, or at least a `renv.lock` file generated with `renv::snapshot()`). In any case, I think the Discussion should better clarify the different roles of software citation (credit, reproducibility, etc) and revise their recommendations accordingly.

Response 9: Great suggestions. We have added some information to the intro and discussion (see responses 7 and 8).

(3) Fig. 2 seems confusing. In panels a and b, the 'No' category appears in black in the bottom. But in other panels it is the 'Yes' category which appears in black in the bottom. I would suggest to revise the figure design to make it more consistent across panels.

Response 10: Indeed. We have changed it now (see response 29).

Other comments

(4) L73-76: "Although container technology such as Docker, which packages the software and its dependencies into a standardized environment, has been suggested as a solution to improving portability and reproducibility, its adoption remains low".

I share the same impression about low adoption of Docker among researchers, but is there any study or numbers to support that statement? For example, how many articles sharing code include a Dockerfile? I do not know of specific studies, but I think it would be useful to cite them if available. Maybe Carl Boettiger may know? (cf. https://doi.org/10.1038/s43586-023-00236-9)

Response 11: Great point. We have searched through several surveys on data- and code-sharing across fields only to find one study, Venkatesh et al. 2022 (https://doi.org/10.1148/ryai.220081; in Radiology AI) showing that of the 73 articles out 218 that shared accessible code, four (6%) shared Docker containers for their models, and for the subset of 24 articles with reproducible code sharing, three (13%) shared Docker containers. We are not aware of any additional studies explicitly estimating the proportion of studies sharing code that use Docker or similar containers and virtual environments. The only additional information we were able to find is that the following study also reports (without providing real evidence) that they are uncommon: Essawy et al. 2020 (https://doi.org/10.1016/j.envsoft.2020.104753). We have updated the text to add Venkatesh et al. 2022 as a reference:

“Although container technology such as Docker, which packages software and its dependencies into a standardized environment, has been suggested as a solution to improve portability and reproducibility (Boettiger 2015; Grüning et al. 2018; Essawy et al. 2020; Trisovic et al. 2022), its adoption remains low (e.g., Venkatesh et al. 2022).”

(5) L83-86: "code should be shared in a permanent repository (e.g., Zenodo) and assigned with an open and permissive licence and a persistent identifier such as a DOI. This is particularly important given the far-from-ideal rates of link persistence found for scientific code in fields such as astrophysics (Allen et al. 2018)".

Sperandii et al. 2024 (https://doi.org/10.1111/jvs.13224) state that "For code, the most frequent reason for inaccessible code was a broken link (69.2%), whereas the use of private repositories was much less common (7.7%)". Perhaps they could be cited here to support that statement.

Response 12: We agree. Reference added.

"*Furthermore, code should be shared in a permanent repository (e.g., Zenodo) and assigned an open and permissive licence and a persistent identifier such as a DOI (Krafczyk et al. 2021; Kim et al. 2022; Jenkins et al. 2023). This is particularly important given the far-from-ideal rates of link persistence found for scientific code in fields such as astrophysics (Allen et al. 2018; Sperandii et al. 2024).*"

(6) L150 and Fig. 2: better explain what 'partially free' software means? Is that a combination of free and propietary software within the same paper?

Response 13: Yes, indeed. We have now clarified this as:

"*and whether the software used was free (i.e., non-proprietary; levels: yes, no, partially; where partially refers to having used both free and non-free software).*"

And also added the following to the caption of Figure 2 (now Figure 3):

"*"Partially free" in panels g and h refers to articles that used both free and non-free software.*"

(7) L193-198: This is fine as is, but I wonder if, for the sake of clarity, these results would rather be reported in a positive framing, e.g. rather than "36.1% (N = 100) of articles published in journals without a code sharing policy did not report the version of all software used", write "63.9% (N = XXX) of articles published in journals without a code sharing policy reported the version of all software used". Sounds less convoluted. Applies also to L210-212.

Response 14: We agree with this suggestion and have changed all this reporting accordingly, including Figure 2 (now Figure 3).

"*Our survey showed that 88.2% of articles (N = 277) published in journals without a code-sharing policy stated the analytical software used (Figure 3a), a value that is almost identical to the 89.9% (N = 311) found for articles published in journals with a code-sharing policy (Culina et al. 2020; Figure 3b). For those stating the statistical software used, 63.9% (N = 177) of the articles published in journals without a code-sharing policy*

reported the version of all software used (Figure 3c), whereas that percentage was 50.2% (N = 156) for articles published in journals with a code-sharing policy (Figure 3d).”

“*For articles stating that they used additional packages, 32.4% (N = 46) of the articles published in journals without a code-sharing policy provided the version of all packages used (Figure 3e), whereas that percentage was 19.5% (N = 40) for articles published in journals with a code-sharing policy (Figure 3f).”*

(8) L213-214: "The mean number of packages used was 2.30 (median = 2.00, range = 1 to 10) in journals without a code sharing policy and 2.41 (median = 2.00, range = 1 to 14) in journals with a code-sharing policy". As far as I understand, this is the number of packages cited in the main text. The actual number of packages used in the analyses is probably quite larger. Please rephrase.

Response 15: Great catch. We have rephrased it accordingly:

"*The mean number of packages reported was 2.30 (median = 2.00, range = 1 to 10) in journals without a code-sharing policy and 2.41 (median = 2.00, range = 1 to 14) in journals with a code-sharing policy."*

(9) L245-246: "several transparency indicators, including, but not limited to, data- and code-sharing are on the rise in ecology (Evans 2016; Culina et al. 2020; Roche et al. 2022a)". Include results from Maitner et al 2024 here too (https://doi.org/10.1002/ece3.70030)?

Response 16: Added:

"*several transparency indicators, including, but not limited to, data- and code-sharing, are on the rise in ecology (Evans 2016; Culina et al. 2020; Roche et al. 2022a; Maitner et al. 2024)*"

References

Boettiger et al. 2015 reference is missing

Response 17: It should have been Boettiger 2015, and we had forgotten to add it to the reference list too. Apologies. We have fixed this now.

Update Ivimey-Cook et al (in prep) to preprint reference?

Response 18: Done.

Review by Veronica Cruz, 06 Feb 2025 16:56

In my review I have evaluated the main points recommended by PCIEcol. In addition, I attach the manuscript in pdf with some minor comments in case the authors want to consider them. I hope the comments are useful and I apologize for my late answer.

Title

Check that the title clearly reflects the content of the article.

It does.

Abstract

Check that the abstract is concise and presents the main findings of the study.

It does.

Introduction

Check that the introduction clearly explains the motivation for the study. Check that the research question/hypothesis/prediction is clearly presented. Check that the introduction builds on relevant recent and past research performed in the field.

To my understading, the introduction explains the motivation of the study and presents clear objectives and hypotheses. However, the results and discussion go beyond these objectives. The novelty of the study is not so clear. The authors cite other articles (even authored by them*) that cover very similar topics. I am not very familiar with the literature in the field but it seems updated and rich.

Response 19: We thank the reviewer for highlighting the effort we put into integrating old and recent literature on this topic, as well as literature from ecology and evolution, and other fields. We thought it was important to do so in this piece, so we appreciate that the reviewer noticed. Thank you also for the helpful review of our preprint.

*The article Ivimey-Cook et al. in prep is cited in the introduction and discussion several times. I find it inappropriate since an article in preparation may end up with very different conclusions. I encourage the authors to upload a preprint of this article if they want to use it to support the manuscript or use it only to support very specific statements of the manuscript.

Response 20: We apologise for this. We were undecided about whether to include this citation or not and decided to do so, since we knew the preprint for Ivimey-Cook et al. would be available by the time our study was being reviewed and thought it would be useful for the reviewer. We have now updated the reference and provided the link to the preprint of that study, which is currently also under review in a journal.

Materials and methods

More generally, check that sufficient details are provided for the methods and analysis to allow replication by other researchers.

I was surprised that given the topic of the article, the authors did not mention in the main text that data management and analysis were done using R and which were the main packages used. I think this is mandatory even though they provide scripts with all the necessary information.

Response 21: We cite both R v.4.3.1 (R Core Team, 2023) and the software used for screening (i.e., Rayyan: Ouzzani et al. 2016). For our analyses, we also used the following packages: stringr, readxl, dplyr, ggplot2, patchwork, and ggpattern. We now also cite ggplot2 in the main manuscript as:

"*Figures were generated using the R package 'ggplot2' v.3.5.1 (Wickham 2016).*"

As reviewer #1 pointed out, we should not expect all packages and dependencies to be cited, but only the fundamental ones, which we agree with (see *Response 6*), thus, we restricted the citation of the other "less significant/auxiliary" packages to the code.

There are several decisions on methodology that seem arbitrary to me, and I think they need further explanation. Although authors followed a methodology that makes results comparable to those in Culina et al. 2020 and thus they refer frequently to this article, some important decisions should be briefly justified in the current article as well. For example, I don't think that separating data in two periods is fulfilling the objectives stated by authors here.

Also, I appreciate the honesty of saying that studies with landscape analysis were discarded because authors "lacked the expertise to understand the analyses and software used.". However, what is the impact of discarding landscape (and molecular analysis as well)? They might be discarding a large number of articles from journals such as Landscape Ecology. If the authors assume that these journals are representative of ecology, why discard certain disciplines? If the scope of the review is missing some parts, it should be clearly stated in results/conclusions/summary.

Response 22: Although we agree discarding landscape papers could be potentially problematic, we do not think it applies to our dataset since it could not have any considerable influence on our results. As we report in our preprint, it only affected two articles, and our final number of articles included was 314:

"*In addition, we excluded two articles from the 2018-2019 subset that performed landscape analyses because we lacked expertise to understand the analyses and software used.*"

Our dataset included 44 articles from the journal Landscape Ecology (2015-2016: 25, 2018-2019: 19), so we consider the room for those two articles from 2018-2019 to have much of an influence on our results to be arguably negligible. In addition, we are clear throughout our preprint that our inferences are made for only non-molecular, so we do not think that focusing on non-molecular and non-landscape (n = 2) articles would bias our results, but simply limit the scope of the study.

Regarding whether "separating data in two periods is fulfilling the objectives stated by authors here", we were interested in trends over time, and given that our search focused on only five years, we used these two periods rather than treating year as a 'continuous' variable with only 5 possible values, as we did in Culina et al. (2020).

Check that the statistical analyses are appropriate.

They are exploratory analysis, and I find them appropriate.

Results

I think the results did not exactly address the objectives. They extensively described the results for journals without code-sharing policies, but they only make a comparison at the end of the first results section (L189) with journals with code-sharing policies. I think that results should be more focused in comparison and maybe report results for journals without code-sharing policies in supplementary materials. They did not state any objective or hypothesis related to different time periods, yet they extensively reported differences between two periods in text and figures and discussed it deeply.

Response 23: We thank the reviewer for this great point. The reason why we are first focusing on journals without policies here and then only presenting results for some of the variables for the subset of journals with code-sharing policies is that those results were already reported in Culina et al. (2020) and we wanted to avoid as much as possible reporting the same results again in this manuscript to avoid duplication. This study is conceived as a follow-up to Culina et al. (2020), and it relies heavily on what was reported in that study. Although this may look a bit suboptimal, we are limited on what we can re-report in this study. To clarify this further, we have reworded the final paragraph of the introduction, which now reads as:

"*Our main goal was to examine whether the implementation of code-sharing policies by journals leads to higher rates of code-sharing and overall reproducibility potential. For that, we assessed the code-sharing and reporting features of 314 articles published in 12 ecological journals without a code-sharing policy, and compared them with those from the comparable sample of 346 articles published in 14 ecological journals with a code-sharing policy from Culina et al. (2020). We predict that ecological journals without a code-sharing*

policy will have lower rates of sharing than journals with a code-sharing policy. However, we do not have a clear expectation of whether the reporting of features associated with higher long-term reproducibility, such as the software used, its versioning and accessibility (free or not), and the location where code is shared, will differ between both sets of journals."

As well as the first sentences of the methodology:

"*Our study design closely matches that of Culina et al. (2020), who surveyed 14 ecological journals with a code-sharing policy from 2015 to 2019. Thus, readers are referred to Culina et al. (2020) for further information on methodology, and importantly, results for journals with a code-sharing policy. In the follow-up study here, we aimed to identify 14 comparable ecological journals without a code-sharing policy for the same period (i.e., 2015-2019).*"

Regarding the time-related hypotheses, we had added the following text to the end of the introduction:

"*Finally, we anticipate that code availability and the reporting of features associated with higher long-term reproducibility will increase over time, regardless of the existence of code-sharing policies, given recent changes in scientific attitudes and norms and the rise of open science (Cao et al. 2023).*"

If possible, evaluate the consistency of raw data and scripts.

All literature reviews or meta-analysis that I know analyze at least the title and abstract of the entire set of articles given by the literature search. I don't think that analyzing a sample of less than 10% of the entire population of the bibliography can make results generalizable. I recommend either expanding the sample of analyzed articles (in journals with and without code-sharing policies) or acknowledging this limitation and including the impact of this decision in the discussion of the results and in conclusions. The subset of articles with available code is very small and could substantially change if another random sample of articles is studied.

Response 24: Although we agree with the observation that most literature reviews or meta-analyses would not follow this approach, it should be noted that ours is neither of those. In addition, the number of articles for which we extracted data in our study (314 without policy + 347 with policy = 661 articles in total) is several times higher than the number of articles included in most meta-analyses in ecology and evolution and beyond (e.g., median = 41 studies in plant ecology: Koricheva and Gurevitch 2014: https://doi.org/10.1111/1365-2745.12224; 44 in evolutionary ecology: Table S6 from Pollo et al. 2024: https://doi.org/10.1111/brv.13117). For meta-research studies like ours, it is customary to randomly sample studies when data extraction is done manually, since manually extracting data from all published studies would not be possible.

In script 003_plotting, the ggplot2 and patchwork libraries need to be explicitly loaded.

Response 25: We suspect that the reason why those two libraries may have had to be loaded explicitly might have been related to the reviewer not having the R package "pacman" installed. To avoid that confusion in the future, we have added the following line of code to all three scripts:

```
##############################################################################
# Packages needed

# if not already installed, run:

#install.packages("pacman")
```

I have some problems with long directory addresses and file names.

Response 26: It is unclear to us what exactly those problems were, but following our instructions in the Readme: "*To run this repository, first download and unzip the entire repository and open the Rproject file: analyses.Rproj. The scripts are named in the order that they should be run, from 001 to 003.*" should not lead to path problems unless the reviewer is hitting the maximum number of characters allowed by Windows, in which case, the reviewer could try setting the repository in, for example, the Desktop, and test if that may fix the problem. We apologise for any inconvenience.

If necessary, and if you can, run the data transformations and statistical analyses and check that you get the same results. In the case of negative results, check that there is a statistical power analysis (or an adequate Bayesian analysis).

I did not check all the results, but the ones I did were consistent.

Response 27: Thank you so much for this!

Inform the recommender and the managing board if you suspect scientific misconduct.

I did not suspect scientific misconduct.

Tables and figures

Check that figures and tables are understandable without reference to the main body of the article. Check that figures and tables have a proper caption.

In Table 1, I don't understand what is "[using data]". Also, here and in other parts of the manuscript, there is no need to say "nonmolecular" articles.

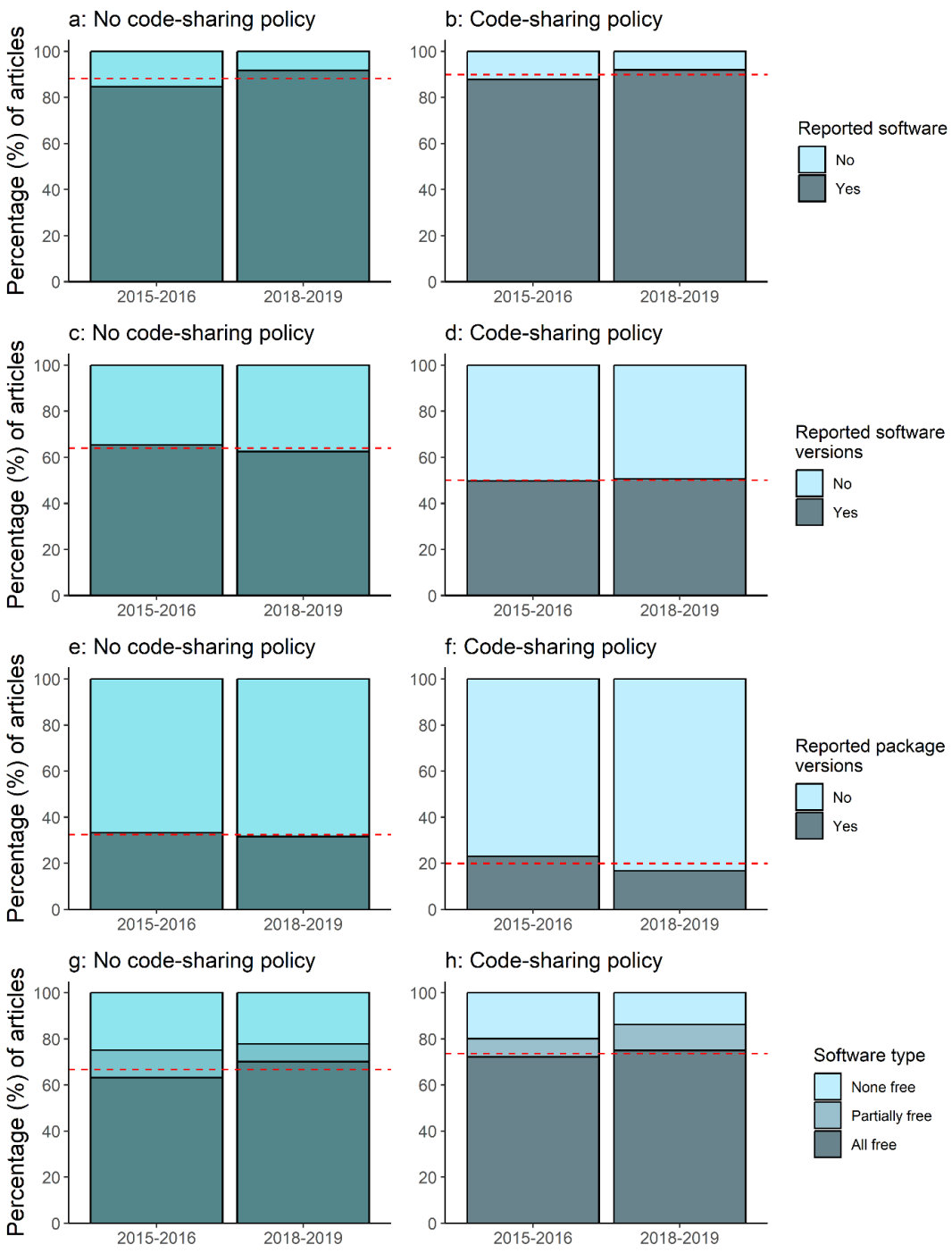
Response 28: That number is calculated taking into account that a few studies in our sample did not contain data because they were simulations. We have added the following clarification to the table caption:

"***Table 1.*** *Code- and data-sharing for 314 nonmolecular articles that conducted statistical analysis or simulations [i.e., did not use data] published between 2015 and 2019 in 12 ecological journals without a code-sharing policy."*

Figure 1 is clear.

Figure 2 can be improved. The dotted pattern is distractive and not needed. Also, it is confusing that sometimes "yes" is white and sometimes is black. I recommend, if possible, to use same color for the desirable output (i.e. reported software, reported versions, free software) all the time. It is also confusing that the caption does not first explain the first panel.

Response 29: These are great points. Thank you very much. We have now changed all figures (and text) to report the "Yes" cases as suggested by both reviewers, and also tried to improve Figure 2, which is now Figure 3 (since we have also included a new figure to the manuscript based on the reviewers' comments). Figure 3 now looks like:



The newly added Figure 2 is:

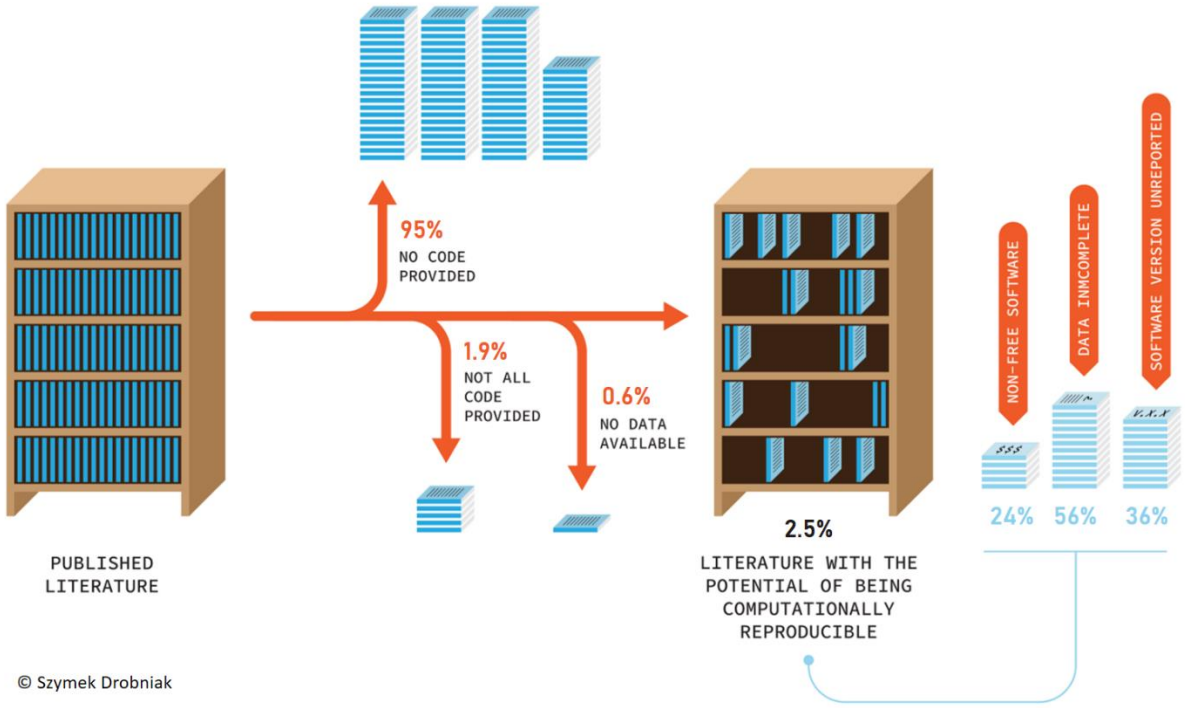


***Figure 2**. Diagram visually representing the computational reproducibility potential of articles published between 2015 and 2019 in ecological journals without a code-sharing policy. The value corresponding to "Data incomplete" used in the diagram (56%) was obtained from Roche et al. (2015); whereas all the remaining values correspond to the survey presented in the current study. Original illustration by Szymek Drobniak.*

Discussion

Check that the conclusions are adequately supported by the results and that the interpretation of the analysis is not overstated.

I think conclusions are adequate, except that the results do not support calling for funders to introduce code- and data-sharing policies.

Response 30: We added this as a recommendation that we think should further increase sharing.

Check that the discussion takes account of relevant recent and past research performed in the field.

I am not very familiar with literature in the field. As a general comment I think that the discussion is biased towards temporal trends in code-sharing rather than the differences between journals with and without code-sharing policies.

Response 31: We agree with the reviewer, but the main reason for this is that few studies have attempted to estimate the effect of policies on sharing, which is why we conducted our study.