The Relationship Between Impact Factor, Discipline, and Data Sharing Policies in Journals

Chris O'Neill, NIEHS
Lucy Evans, Duke University
David Martin, Duke University
Sophia Williams, NIEHS
Sarah Green, OHSU-PSU School of Public Health, Oregon Health & Science University, Portland, OR, United States
Thomas Harris, Oregon Clinical & Translational Research Institute, Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, OR
Matthew Lewis, OHSU Library, Oregon Health & Science University, Portland, OR, United States

Abstract

Data sharing is considered a crucial aspect of the research process, facilitating transparency and allowing for replication and re-analysis, with the ultimate goal of advancing science. Publishers are well positioned to promote sharing of research results. While there is an increasing trend towards requiring data sharing, not all journals require data be shared at the time of publication. In this study, we extended previous work to analyze the data sharing policies of 447 journals across several scientific disciplines, including biology, clinical sciences, mathematics, physics, and social sciences. Our results showed that only a small percentage of journals require data sharing as a condition of publication, and that this varies across disciplines and Impact Factors. Both Impact Factor and discipline are associated the presence of a data sharing policy. Our results suggest that journals with higher Impact Factors are more likely to have data sharing policies; use shared data in peer review; require deposit of specific data types into publicly available data banks; and refer to reproducibility as a rationale for sharing data. Biological science journals are more likely than social science and mathematics journals to require data sharing.

Introduction

Sharing research results, data, and other intellectual products is an established and essential part of the scientific ethos (Resnik 1998). The eminent sociologist of science Robert Merton described a norm, known as communalism, which obliges scientists to share the results of research. Communalism holds that the findings of science belong to the research community because they are the products of collaboration (Merton 1973). Collaboration, in turn, promotes scientific progress by allowing researchers to pool their resources and work toward common goals (Fischer and Zigmond 2010).

Data sharing also plays a key role in allowing researchers to evaluate, re-analyze and reproduce each other's work (Fischer and Zigmond 2010, Bauchner et al. 2016). Recently, scientists and others have raised concerns about the reproducibility of published research (Arrowsmith 2001, Ioannidis 2005, Prinz et al. 2011, McNutt 2014, Open Science Collaboration 2015, Begley and Ioannidis 2015, Baker 2016). In response, funding agencies, professional societies, and journals

have developed or refined policies to promote transparency, including the creation of data sharing policies (Collins and Tabak 2014, McNutt 2014, Vasilevsky et al. 2017, Stuart et al. 2018). These policies encourage or require researchers to share the information needed to reproduce research results, such as supporting data, protocols, study designs, methods, and computer code (Kaye et al. 2009, Stuart et al. 2018, Research Councils UK. 2011, NIH 2016a, NIH 2016b, Holdren 2012). Although most scientists agree with the importance of data sharing, several barriers and practical and ethical challenges are sometimes raised. These include issues such as protecting human subjects, concerns about "scooping," intellectual property considerations, proprietary information, and national security issues. In addition, the time and costs associated with sharing data can be significant (Resnik 2009, Stodden et al. 2013, Shamoo and Resnik 2015, Longo and Draze, 2016, Stuart et al. 2018, LeClere 2010, Savage & Vickers 2009).

Journals play an important role in promoting data sharing. Several previous studies have examined journals' data sharing policies. (Stodden et al. 2013, Vasilevsky et al. 2017, Piwowar et al. 2008, McCain 1995, Barbui 2016). Stodden et al. (2013) developed and tested a framework for classifying the data and code sharing policies in computational and mathematical biology, statistics and probability, and multidisciplinary science journals for the years 2011 and 2012. Their results showed that 38% of journals they evaluated had a data sharing policy and 22% had a code sharing policy. A journal's Impact Factor was significantly associated with having a policy; journals with higher Impact Factors were more likely to have data sharing policies or impose significant requirements on data sharing (Stodden et al. 2013). Vasilevsky et al. (2017) refined Stodden et al.'s coding framework to evaluate additional aspects of data sharing requirements and expanded the subject scope and number of evaluated journals. They found that 11.9% of journals required data sharing as a condition of publication, 9.1% required data sharing but did not state whether it would affect publication decisions, 23.3% encouraged or addressed data sharing, 9.1% mentioned data sharing indirectly, 14.8% of journals addressed protein, proteomic, and/or genomic data sharing, and 31.8% did not mention data sharing (Vasilevsky et al. 2017). Like Stodden et al. (2013), the authors also found that having a data sharing policy was significantly associated with the Journal Impact Factor. The Stodden et al. (2013) and Vasilevsky et al. (2017) studies focused on biomedical or computational journals. Moreover, Vasilevsky et al.'s study was limited to journals in the top quartile by Impact Factor or publication volume.

Given the practical and ethical challenges related to data sharing discussed above, one would expect there to be some diversity in data sharing norms, attitudes, and policies across different areas of science, and evidence from the published literature provides support for this hypothesis (Kaye et al. 2009, Sorrano et al. 2015, Stuart et al. 2018). In a survey of almost 8,000 researchers, Stuart et al. (2018) found that attitudes towards data sharing varied across disciplines such as biological, clinical, and physical sciences. The goal of our research was to build upon this prior work to examine the relationships between the presence and characteristics of data sharing policies in various journal subject categories and by Journal Impact Factor. We hypothesized that Journal Impact Factor and type of science would affect the presence and characteristics of data sharing policies. Towards this end, we examined a large sample of journals from varied scientific disciplines with a wide range of Impact Factors.

Methods

We drew our sample from the 2016 edition of Thompson-Reuters' Journal Citation Reports, which included 13,401 journals. Unlike Vasilevsky et al. (2017), we did not limit our sample to high Impact Factor journals. To ensure that our sample contained journals from a variety of disciplines, we drew a stratified random sample of journals. For our stratification, we categorized the journals listed in Journal Citation Reports as belonging to the following disciplines:

- **Biological sciences** (2263 journals); i.e. journals that publish research in biology, biotechnology, bioengineering, ecology, agriculture.
- **Clinical sciences** (4132 journals), i.e. journals that aim to publish clinical research or have a substantial proportion of articles that report the results of clinical research; includes medical journals and social science journals with a clinical focus.
- **Mathematical sciences** (1280 journals), i.e. mathematics, statistics, and computer science; also includes mathematically oriented journals in biology, physics, and social science.
- **Physical science** (2709 journals), including engineering; excluding bioengineering.
- **Social sciences** (2953 journals), including humanities.
- **Multidisciplinary science journals** (64 journals), e.g. *Science*, *Nature*, etc.

We initially grouped journals according to these different categories based on the names and descriptions found in Journal Citation Reports and then reclassified some of the journals during the coding process.
We eliminated multidisciplinary science journals to facilitate more consistent disciplinary comparisons. Using the website Random.org, we drew a random sample of 125 journals form each journal category, and eliminated 19 journals because they were review journals or duplications. Additionally, we removed 13 journals because they were not in English (8), , were books (3) or did not post their policies on the internet (2). That left us with 593 journals to code from five different areas of science. Ultimately, we coded only 447 journals due to limited resources. The breakdown was: 18.1% biological sciences, 18.8% clinical sciences, 21.7% mathematical sciences, 19.9% physical sciences, and 21.5% social sciences. The mean Impact Factor was 2.0, with a range of 0.13 to 29.5 and a standard deviation of 2.60. 88.1% of journals were available by subscription and 11.9% were open access (See Table 1).

The coders (RL and MM, interns) worked independently. Each reviewed the data sharing policies documented on journal websites and coded the policies based on a framework adapted from Stodden et al. (2013) and Vasilevsky et al. (2017). The coders limited their evaluations to information being communicated to manuscript submitting authors, and did not consider external sources of information, such as footnoted links to additional web pages, unless authors were specifically instructed to review this information in order to understand or comply with a journal's data sharing policy. The coders resolved any conflicts after inter-rater agreement was statistically assessed. The coders were able to resolve their disagreements by rereading the policies and consulting with a third party, DR. The average agreement was 86.8% and agreement was 90% or above for 13 of 24 variables assessed. (Inter-rater agreement data analyses are available upon request.)

The coding framework consisted of the following categories:

Data sharing policy (note: categories are not mutually exclusive)

Data sharing required as condition of publication
Data sharing required but no explicit statement regarding effect on publication/editorial decision-making
Data sharing explicitly encouraged/addressed but not required
Shared data will be used in peer review
Data sharing mentioned indirectly
Only protein, proteomic, and/or genomic data sharing addressed
Only sharing of computer code addressed
No mention of data sharing

Specific types of data sharing

Sharing of protein, proteomic, genetic, or genomic data sharing required with deposit to specific data banks
Sharing of computer codes or models with deposit to data banks
Sharing of clinical trial data with deposit to data banks

Recommended sharing method

Public online repository
Journal hosted
By reader request to authors
Multiple methods equally recommended
Unspecified

If data is journal hosted

Journal will host regardless of size
Journal has data hosting file/s size limit
Unspecified

Issues concerning copyright/licensing of data

Explicitly stated or mentioned
No mention

Data archival/retention policy

Explicitly stated
No mention

Reproducibility or analogous concepts noted as purpose of data policy

Explicitly stated
No mention

The coders also confirmed the original classification of the journal's subject category (i.e. biological, clinical, etc.) made during the stratification process. 8 journals were re-classified based on careful inspection of their aims and scope and the type of articles published. Each journal's Impact Factor and access model (open access or subscription) were also recorded. We used the statistical software R (version 3.5.0) in our data analyses. Logistic regression models (Table 2) were used to assess the association between Impact Factor, discipline, and the presence and type of data sharing policy. We included both Impact Factor (continuous variable) and discipline (categorical variable) as predictors and each data sharing policy as the response variable in turn in the model. To ensure reasonable sample sizes we only performed logistic regression on the policies with >20 "Yes" responses. Likelihood ratio tests were performed to evaluate the global association of type of science.

Results

Of the 447 journals evaluated, only 12 journals (2.7%) required data sharing as a condition of publication. 35 (7.8%) required data sharing but did not explicitly state the effect on publication. A total of 181(40.5%) encouraged or addressed data sharing but did not require it. 25 (5.6%) stated that shared data would be evaluated during peer review. 24 (5.4%) indirectly mentioned data sharing, while 12 (2.7%) only addressed proteomic and/or genomic data sharing; 7 (1.6%) only addressed computer code sharing. 43.6% of the journals (195) did not mention data sharing at all (Figure 1). Note, scoring for each category was not mutually exclusive.

With respect to specific types of data sharing, such as the mention of specific repositories, we found that a small percentage of journals specifically required data deposits into public repositories. A total of 44 (9.8%) of journals required the deposit of protein, proteomic, genetic, or genomic data into specific data banks. 90 (20.1%) journals required sharing of computer codes or models with deposit into data banks, including. A small number of journals, 11 (2.1%) required deposit of clinical trial data into clinical data banks (Figure 2).

While many data repositories exist, providing recommendations for data deposition or hosting the data at the journal itself may reduce barriers to data sharing and improve downstream user access. We therefore analyzed the recommended sharing method for each journal. 189 (42.3%) of journals recommend data sharing via a public online repository, 19 (4.5%) recommend data sharing through a journal hosted repository, 8 (1.8%) recommend data sharing by a reader request to the authors, 32 (7.2%) recommend multiple methods of sharing data, and 200 (44.7%) did not specify the method of data sharing. Among the 19 journals that offered to host data, 2 journals (10.5%) mentioned they would host the data regardless of the size of the dataset, 2 (10.5%) placed limits on the size of the dataset or number of files, and 15 (78.9%) made no mention of hosting restrictions (Figure 2). Additionally, data can be licensed it ways that facilitate or hinder reuse and attribution by others. Our results showed that only 9 (2%) journals explicitly stated or mentioned issues related to copyright or licensing of data, and 4 (0.9%) explicitly stated a data archiving or retention policy. Finally, a driving factor for data sharing is to facilitate research reproducibility; therefore, we analyzed policies for the mention of reproducibility or analogous concepts as the purpose of data sharing. Our results showed that

approximately one quarter of the journals (106 (23.7%)) referred to reproducibility or related concepts as justification for their data sharing policy (Figure 2).

Impact Factor was significantly associated with the several aspects of data sharing. Journals with a high Impact Factor were more likely to use data in the peer review process with an odds ratio (OR) of 1.17 (95% confidence interval (CI)= (1.06, 1.28)) per unit increase in Impact Factor; are less likely to "not mention data sharing policy" (OR=0.76 (0.65, 0.89)); more likely to require deposit of protein, proteomic, genetic, and genomic data into a public data bank (OR=1.2 (1.05, 1.37)) and computer code deposit (OR=1.09 (1, 1.18)); are more likely to recommend sharing on public online repository (OR=1.24 (1.1, 1.41)); less likely to "not recommend a sharing mechanism" (OR=0.75 (0.59, 0.83)); and more likely to mention reproducibility or analogous concepts (OR=1.26 (1.11, 1.42)) (Table 2).

We also observed differences in data sharing policies across five scientific sub-disciplines. For all comparisons we used journals in biological sciences as the reference group, which allowed us to build upon and compare our findings to previous studies. Our results showed that both social science journals and mathematical science journals are less likely than biological science journals to require data sharing (with no impact on the publication decision) (OR=0.12 (0.03, 0.53) and 0.17 (0.05, 0.63), respectively). As for recommended data sharing mechanism, mathematical science journals are less likely to recommend a public online repository (OR=0.54, 95% CI= (0.29, 1)) or recommend multiple methods (OR=0.11, 95% CI= (0.01, 0.94)). Not surprisingly, comparing to journals in biological sciences, journals in clinical sciences, physical sciences, and social sciences from our analysis were less likely to require deposit of protein, proteomic, genetic, and genomic data into a public data bank. The 4 degrees-of-freedom likelihood ratio tests of disciplines showed significant association with several data sharing policy characteristics: data sharing required but no effect on publication described (P=0.0024), required deposit of protein, proteomic, genetic, and genomic data into a public data bank (P=7.8 x 10^{-10}), and multiple methods recommended for data sharing (P=0.02). Discipline is also significantly associated with journal access type (P=0.04) with social science journals being less likely to be open access (OR=0.24 (0.08, 0.7)) (Table 2).

Discussion

The benefits of data sharing have been well documented as a means to make science more transparent and reproducible and to advance discovery, as highlighted in cases such as the Human Genome Project (https://www.genome.gov/12011238/an-overview-of-the-human-genome-project/) and the Framingham Heart Study (https://www.framinghamheartstudy.org/). As most scientific findings are communicated via published manuscripts, publishers are uniquely positioned to promote and require data sharing as an outcome of research. However, our analysis of a large subset of journals across a range of Impact Factors and scientific sub-disciplines validate and extend previous findings that most journals do not require data sharing alongside publications (Stodden et al 2013, Vasilevsky et al 2017, Piwowar et al 2008, McCain, 1995, Barbui, 2016).

We found that the presence and quality of data sharing policies varied across journals in various disciplines and Impact Factors. Journals with higher Impact Factors were more likely to have

data sharing policies; use shared data in peer review; require deposit of protein, proteomic, genetic, and genomic data into publicly available data banks; recommend deposit of data onto shared repositories; and refer to reproducibility as a rationale for sharing data. Biological science journals in our subset were more likely than social science and mathematics journals to require data sharing (with no effect on publication) and were, unsurprisingly, more likely to require deposit of protein, proteomic, genetic, and genomic data into a public data bank than all other types of journals. Biological science journals are also more likely than mathematics journals to recommend deposit of data into a shared repository.

In comparing our results with the previously published study by Vasilevsky et al. (2017), we note some key differences and similarities. With respect to the journals analyzed, the initial study was limited to the top quartiles by numbers of citable items (published papers) and Impact Factor, while this study was not. The previous study sample had a higher mean and median Impact Factor compared to the sample analyzed in this study (original study, mean=5.42, median=4.16; current study: mean=2.22, median 1.75). In the previous study, the authors found a correlation between the number of citable items and the presence of a data sharing policy. The mean and median number of citatable items in the previous study's sample were higher than the journals in this sample (original study, mean=414, median=123; this study, mean=236, median=82). The overall and domain specific result differences between our results and Vasilevsky et al. (2017) can be attributed to sample differences and support both studies' conclusions related to the influence of Impact Factor on the presence of a data sharing policy

The observation that journal Impact Factor is associated with data sharing policy development is not without precedent. AAlsheikh-Ali et al. showed that 44 out of 50 (88%) of the highest Impact Factor journals had a statement about data sharing in their instructions to authors (2011). Additionally, previous studies have shown that Impact Factor is associated with the development of misconduct (Resnik et al. 2010), dual use (Resnik et al. 2011), and retraction (Resnik et al. 2015) policies. Impact Factor may affect policy development because journals with higher Impact Factors tend to receive more attention and scrutiny from scientists and the media, so they are more likely to be asked and feel pressure to create policies that address issues in science and the research literature. Journals with higher Impact Factors are also likely to have more resources (e.g. editorial board member, editorial staff, money) that are needed to develop and implement policies.

Associations between discipline and data sharing policies are also to be expected, since different scientific disciplines have different needs and concerns related to data sharing (Stuart et al. 2018, Taichman et al, 2016, Borgman, 2012, Wallis et al 2013, Tenopir et al 2015). Disciplines in the mathematical, physical, and social sciences, for example, rarely publish research in genomics or proteomics, so they have little need for policies requiring the sharing of proteomic or genomic data. While there are increasing demands for clinical data to be shared (Sommer 2010, Taichman et al., 2017), so one might expect that clinical science journals would be more likely to have a data sharing policy, but we did not observe this association.

Reproducibility and reuse should be important in any field of science, so it is somewhat surprising that biological science journals are more likely to refer to this concept in data sharing policies than journals from other types of science, especially since reproducibility problems have

emerged in the clinical and social sciences (Collins and Tabal 2014, Open Science Collaboration 2015). However biological and health sciences research has received considerable attention from funders regarding reproducibility issues. It remains to be seen whether journals outside of biology will start to place more emphasis on reproducibility as they deal with problems related to adherence to this important scientific norm.

This study confirms previous reports and extends our understanding of journal data sharing policies into other scientific sub-disciplines. However, this study was still limited to five journal subject categories. Data collection based on further refinement of these categories could yield valuable information. We encourage other researchers to conduct studies of journal data policies that focus on narrower categories of science, such as environmental science, psychological science, pharmaceutical science, etc. This data was also captured during a fixed point in time, but publication requirements and data sharing policies are not static. New policies may be implemented or changed over time, which will affect this analysis. Additionally, the inter-curator agreement was high, but still below 100 %, suggesting that some of the policies were ambiguous and subject to interpretation.

In conclusion, while scientific journals have made considerable progress in the last decade with respect to the implementation of data sharing policies, improvements can still be made. yet. Journals that are interested in promoting transparency, reproducibility, and reuse should consider requiring data sharing via clear and specific policies that facilitate these outcomes.

Acknowledgements
This research was supported by the Intramural Program of the National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH). It does not represent the views of the NIEHS, NIH, or U.S. government.

References

1. Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JPA (2011) Public Availability of Published Research Data in High-Impact Journals. PLoS ONE 6(9): e24357. https://doi.org/10.1371/journal.pone.0024357
2. Arrowsmith J. 2011. Trial watch: Phase II failures: 2008-2010. Nature Reviews Drug Discovery10(5):328-329.
3. Baker, M (2016) 1,500 scientists lift the lid on reproducibility. Nature 533:452-454.
4. Barbui C. 2016. Sharing all types of clinical data and harmonizing journal standards. BMC Medicine 14:63
5. Bauchner H, Golub RM, Fontanarosa PB. 2016. Data sharing: an ethical and scientific imperative. Journal of the American Medical Association 315(12):1238-1240.
6. Begley CG and Ioannidis JPA (2015) Reproducibility in Science: Improving the Standard for Basic and Preclinical Research. Circulation Research. 2015;116:116-126
7. Borgman CL (2012). The conundrum of sharing research data. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, 63(6):1059–1078, 2012
8. Collins FS, Tabak LA. 2014. Policy: NIH plans to enhance reproducibility. Nature 505(7485):612-613.

9. Fischer BA, Zigmond MJ. 2010.The essential nature of sharing in science. Science and Engineering Ethics 16(4):783-799.
10. Holdren JP. 2012. Increasing access to the results of federally funded scientific research. 2012. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
11. Ioannidis JP. 2005. Why most published research findings are false. PLoS Medicine 2:696–701.
12. Kaye J, Heeney C, Hawkins N, de Vries J, Boddington P. 2009. Data sharing in genomics--re-shaping scientific practice. Nature Reviews Genetics 10(5):331-335.
13. LeClere (2010) LeClere F. Too many researchers are reluctant to share their data. 2010. The Chronicle of Higher Education. https://www.chronicle.com/article/Too-Many-Researchers-Are/123749
14. Longo & Drazen (2016) Longo DL, Drazen JM. Data sharing. New England Journal of Medicine. 2016;374:276–277. doi: 10.1056/NEJMe1516564.
15. McCain KW. 1995. Mandating sharing: journal policies in the natural sciences. Science Communication 16:403-431
16. McNutt M. 2014. Reproducibility. Science 343(6168):229.
17. Merton RK. 1973. The Sociology of Science. Chicago, IL: Chicago University Press.
18. NIH (2016a) NIH National institutes of health genomic data sharing policy. 2016a. [20 October 2016]. https://gds.nih.gov/03policy2.html .
19. NIH (2016b) NIH Principles and Guidelines for Reporting Preclinical Research — National Institutes of Health (NIH) 2016b. https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research. [20 October 2016]. https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research
20. Open Science Collaboration. 2015. Psychology. Estimating the reproducibility of psychological science. Science 349(6251):aac4716.
21. Prinz F, Schlange T, Asadullah K. 2011. Believe it or not: how much can we rely on published data on potential drug targets? Nature Reviews Drug Discovery 10(9):712.
22. Research Councils UK (2011) Research Councils UK RCUK common principles on data policy—research councils UK. 2011. http://www.rcuk.ac.uk/research/datapolicy/ [25 October 2016]. http://www.rcuk.ac.uk/research/datapolicy/
23. Resnik DB. 1998. The Ethics of Science. New York, NY: Routledge.
24. Resnik DB. 2009. Playing Politics with Science: Balancing Scientific Independence and Government Oversight. New York, NY: Oxford University Press.
25. Resnik DB, Barner DD, Dinse GE. Dual-use review policies of biomedical research journals. Biosecurity and Bioterrorism 2011; 9(1): 49-54.
26. Resnik DB, Patrone D, Peddada S. 2010. Research misconduct policies of social science journals and Impact Factor. Accountability in Research 17(2): 79-84.
27. Resnik DB, Wager E, Kissling GE. 2015. Retraction policies of top scientific journals ranked by Impact Factor. Journal of Medical Librarian Association 103(3):136-139.
28. Savage & Vickers (2009). Empirical study of data sharing by authors publishing in PLoS journals. PLOS ONE. 2009;4(9):e7078 doi: 10.1371/journal.pone.0007078.
29. Sommer J (2010). The delay in sharing research data is costing lives. Nature Medicine volume 16, page 744 (2010)

30. Sorrano PA, Cheruvelil KS, Elliott KC, Montgomery GM. 2015. It's good to share: why environmental scientists' ethics are out of date. BioScience 65(1):69–73.
31. Stodden V, Guo P, Ma Z. 2013. Toward reproducible computational research: an empirical analysis of data and code policy adoption by journals. PLOS ONE 8(6):e67111
32. Stuart D, Baynes G, Hrynaszkiewicz I, Allin K, Penny D, Lucraft M, Astell M. 2018. White paper: Practical Challenges for Researchers in Data Sharing. Available at: https://doi.org/10.6084/m9.figshare.5975011
33. Taichman TB, et al (2016). Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors. Ann Intern Med. 2017;167(1):63-65.
34. Tenopir C, Dalton ED, Allard S, Frame M, Pjesivac I, Birch B, et al. (2015) Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. PLoS ONE 10(8): e0134826. https://doi.org/10.1371/journal.pone.0134826
35. Vasilevsky NA, Minnier J, Haendel MA, Champieux RE. Reproducible and reusable research: are journal data sharing policies meeting the mark? PeerJ 2017 Apr 25;5:e3208.
36. Wallis JC, Rolando E, Borgman CL (2013) If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. PLoS ONE 8(7): e67332. https://doi.org/10.1371/journal.pone.0067332

Table 1: Journal Summary Data

Journal Categories	Count	Percent (%)
Biological Sciences	81	18.1
Clinical Sciences	84	18.8
Mathematical Sciences	97	21.7
Physical Sciences	89	19.9
Social Sciences	96	21.5
Total	447	

Impact Factor*

Mean 2.0; Median: 1.44; Range: 0.13 to 29.5; Standard Deviation: 2.60

*445 journals; 2 had no Impact Factor available

Journal access	Count	Percent (%)
Open access	394	11.9
Subscription	53	88.1

Figure 1: ~~Type of Data Sharing Policy.~~ Percentage of journals per each data sharing mark. Each bar shows the percentage of all journals for each data sharing mark. Each journal was coded with a data sharing mark by two independent curators. The categories were not mutually exclusive.

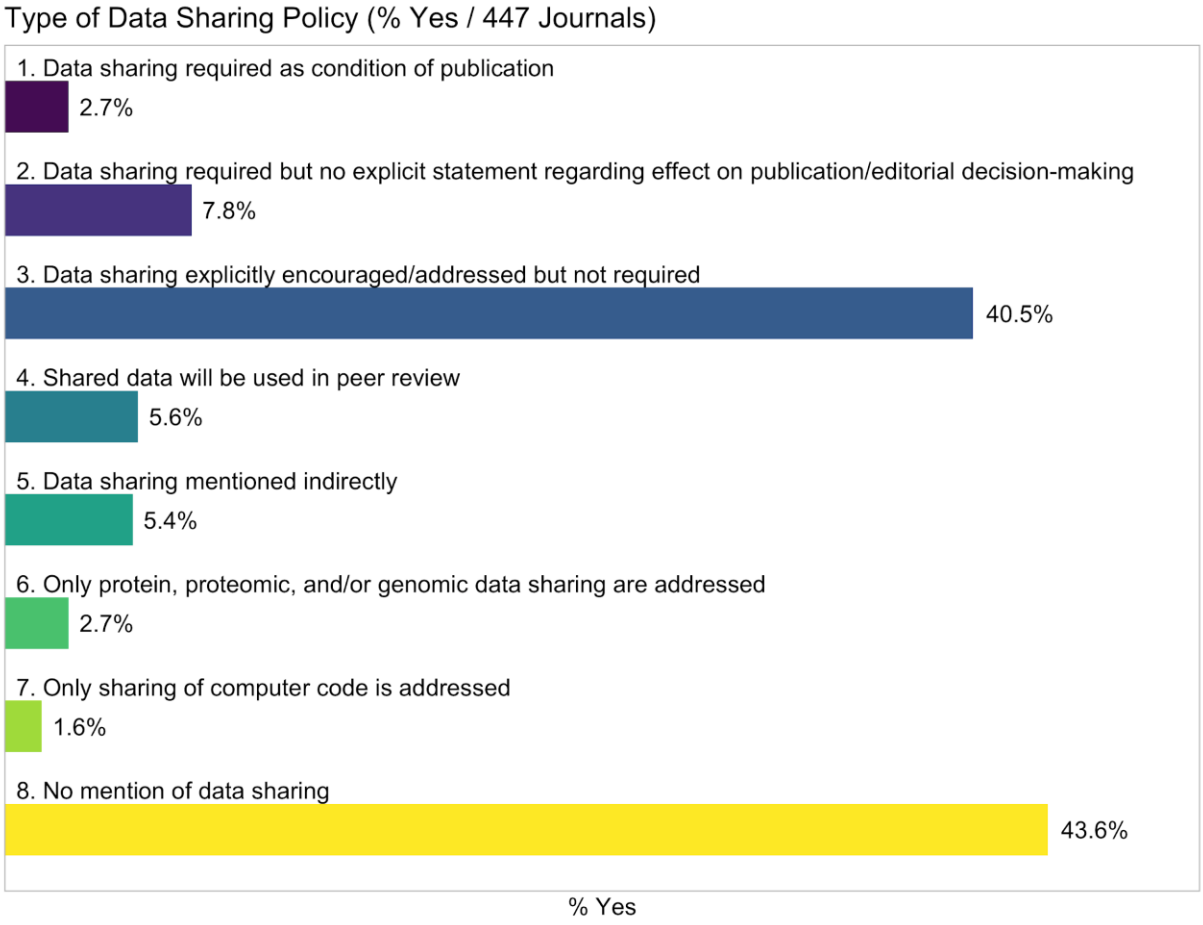

Figure 2: Specific Types of Data Sharing . Percentage of journals per each specific type of data sharing mark. Each bar shows the percentage of all journals for each specific type of data sharing mark, as indicated below. Each journal was coded with a data sharing mark by two independent curators. The categories were not mutually exclusive.

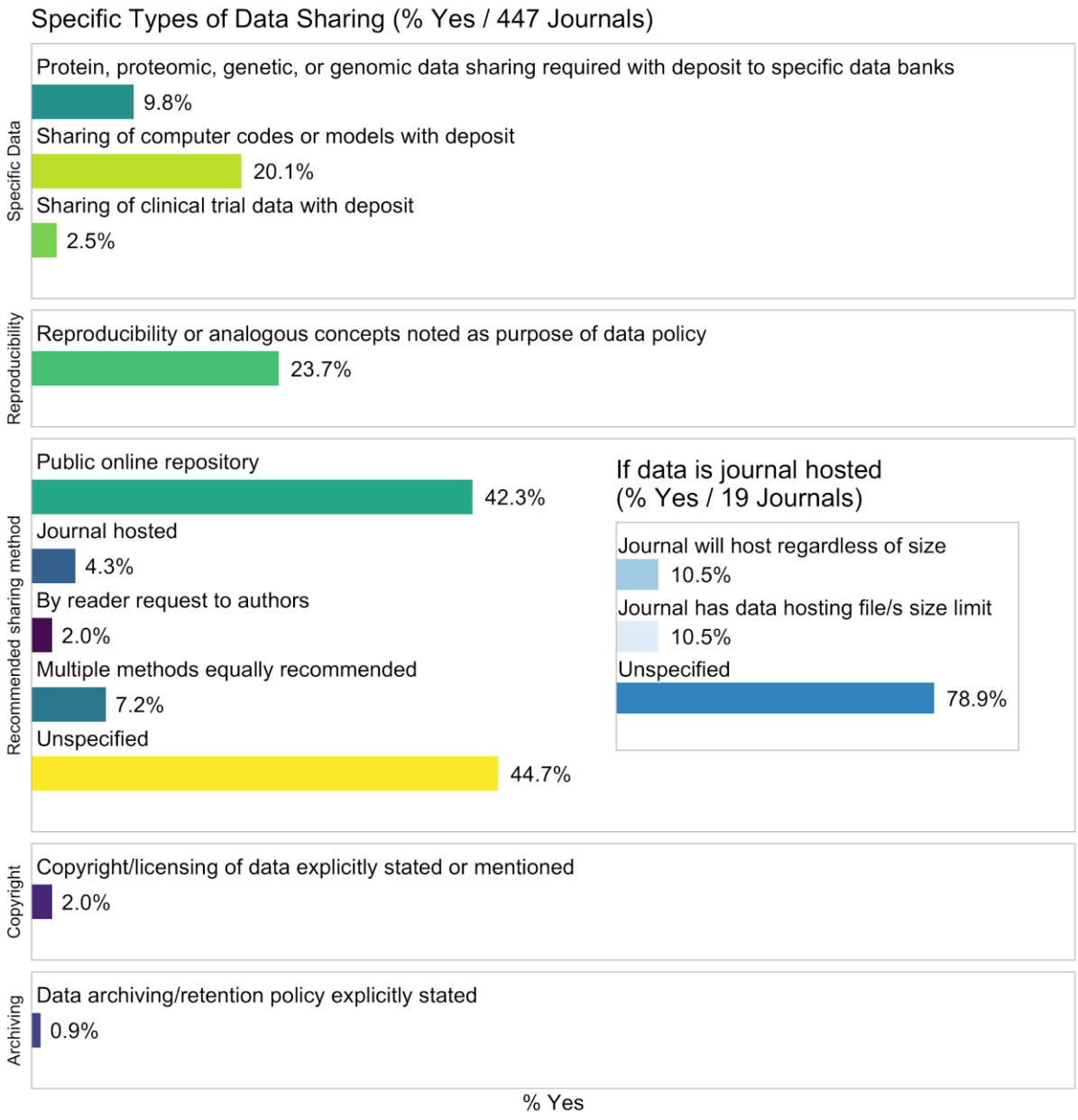

Table 2: Results of logistic regressions

Data sharing policy (Binary dependent variable)	Variable[1]	OR (95% CI)	P (Wald test)	P (Likelihood Ratio Test of Type of Science)
Data sharing required but no explicit statement regarding effect on publication/editorial decision-making	Impact Factor	1.04 (0.95, 1.15)	0.4	
	Type of Science: C	0.70 (0.28, 1.70)	0.4	
	Type of Science: M	0.17 (0.05, 0.63)	**0.0077**	
	Type of Science: P	0.46 (0.17, 1.22)	0.1	
	Type of Science: S	0.12 (0.03, 0.53)	**0.0055**	**0.0024**
Data sharing explicitly encouraged/addressed but not required	Impact Factor	1.07 (0.99, 1.17)	0.1	
	Type of Science: C	1.06 (0.57, 1.99)	0.8	
	Type of Science: M	0.85 (0.46, 1.57)	0.6	
	Type of Science: P	1.13 (0.61, 2.08)	0.7	
	Type of Science: S	1.04 (0.57, 1.91)	0.9	0.9
Shared data will be used in peer review	Impact Factor	1.17 (1.06, 1.28)	**0.001**	
	Type of Science: C	0.88 (0.19, 4.01)	0.9	
	Type of Science: M	1.00 (0.23, 4.40)	1	
	Type of Science: P	1.93 (0.51, 7.35)	0.3	
	Type of Science: S	1.36 (0.33, 5.60)	0.7	0.8
Data sharing mentioned indirectly	Impact Factor	1.01 (0.86, 1.19)	0.9	
	Type of Science: C	5.08 (0.58, 44.57)	0.1	
	Type of Science: M	7.26 (0.89, 59.53)	0.1	
	Type of Science: P	2.84 (0.29, 27.87)	0.4	
	Type of Science: S	6.37 (0.76, 53.12)	0.1	0.1
No mentioning of data sharing	Impact Factor	0.76 (0.65, 0.89)	**0.0005**	
	Type of Science: C	1.21 (0.62, 2.34)	0.6	
	Type of Science: M	1.74 (0.94, 3.24)	0.1	
	Type of Science: P	1.60 (0.85, 3.01)	0.1	
	Type of Science: S	1.55 (0.83, 2.88)	0.2	0.4
Protein, proteomic, genetic, or genomic data sharing required with deposit to specific data banks	Impact Factor	1.20 (1.05, 1.37)	**0.0074**	
	Type of Science: C	0.46 (0.21, 1.01)	0.05	
	Type of Science: M	--	1	
	Type of Science: P	0.15 (0.05, 0.44)	**0.0005**	
	Type of Science: S	0.09 (0.03, 0.33)	**0.0002**	**7.8×10^{-10}**
Sharing of computer codes or models with deposit	Impact Factor	1.09 (1.00, 1.18)	**0.0439**	
	Type of Science: C	1.77 (0.86, 3.66)	0.1	
	Type of Science: M	0.85 (0.39, 1.84)	0.7	
	Type of Science: P	1.00 (0.46, 2.16)	1	

	Type of Science: S	0.81 (0.37, 1.77)	0.6	0.2
Recommended sharing method: Public online repository	Impact Factor	1.24 (1.10, 1.41)	**0.0008**	
	Type of Science: C	0.64 (0.34, 1.20)	0.2	
	Type of Science: M	0.54 (0.29, 1.00)	0.05	
	Type of Science: P	0.72 (0.39, 1.34)	0.3	
	Type of Science: S	0.67 (0.37, 1.23)	0.2	0.4
Recommended sharing method: Multiple methods recommended	Impact Factor	1.04 (0.94, 1.15)	0.4	
	Type of Science: C	1.58 (0.58, 4.30)	0.4	
	Type of Science: M	0.11 (0.01, 0.94)	**0.044**	
	Type of Science: P	0.93 (0.31, 2.78)	0.9	
	Type of Science: S	0.73 (0.23, 2.28)	0.6	**0.02**
Recommended sharing method: Unspecified	Impact Factor	0.75 (0.59, 0.83)	**0.00003**	
	Type of Science: C	1.26 (0.65, 2.43)	0.5	
	Type of Science: M	1.51 (0.82, 2.81)	0.2	
	Type of Science: P	1.42 (0.75, 2.66)	0.3	
	Type of Science: S	1.41 (0.76, 2.62)	0.3	0.7
Reproducibility or analogous concepts noted as purpose of data policy	Impact Factor	1.26 (1.11, 1.42)	**0.0002**	
	Type of Science: C	1.31 (0.66, 2.61)	0.4	
	Type of Science: M	0.76 (0.37, 1.55)	0.5	
	Type of Science: P	0.65 (0.31, 1.36)	0.3	
	Type of Science: S	0.64 (0.31, 1.34)	0.2	0.2
Journal access: Open Access	Impact Factor	0.84 (0.66, 1.06)	0.1	
	Type of Science: C	0.82 (0.34, 1.95)	0.6	
	Type of Science: M	0.45 (0.18, 1.11)	0.1	
	Type of Science: P	0.80 (0.35, 1.84)	0.6	
	Type of Science: S	0.24 (0.08, 0.70)	**0.0093**	**0.04**

1: Journal catetories: C = clinical sciences, M = mathematical sciences, P = physical sciences, S = social sciences. The reference category is B: biological sciences.