

Correspondence

fiddle: a tool to combat publication bias by getting research out of the file drawer and into the scientific community

 René Bernard^{1,2,*},  Tracey L. Weissgerber^{2,*},  Evgeny Bobrov²,  Stacey J. Winham³,  Ulrich Dirnagl^{1,2} and  Nico Riedel²

¹NeuroCure Cluster of Excellence, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany; ²QUEST Center for Transforming Biomedical Research, Berlin Institute of Health (BIH), Berlin, Germany; ³Division of Biomedical Statistics and Informatics, Mayo Clinic Rochester, MN, U.S.A.

Correspondence: René Bernard (rene.bernard@charite.de)



Statistically significant findings are more likely to be published than non-significant or null findings, leaving scientists and healthcare personnel to make decisions based on distorted scientific evidence. Continuously expanding ‘file drawers’ of unpublished data from well-designed experiments waste resources creates problems for researchers, the scientific community and the public. There is limited awareness of the negative impact that publication bias and selective reporting have on the scientific literature. Alternative publication formats have recently been introduced that make it easier to publish research that is difficult to publish in traditional peer reviewed journals. These include micropublications, data repositories, data journals, preprints, publishing platforms, and journals focusing on null or neutral results. While these alternative formats have the potential to reduce publication bias, many scientists are unaware that these formats exist and don’t know how to use them. Our open source **file drawer data liberation effort** (*fiddle*) tool (RRID:SCR_017327 available at: <http://s-quest.bihealth.org/fiddle/>) is a match-making Shiny app designed to help biomedical researchers to identify the most appropriate publication format for their data. Users can search for a publication format that meets their needs, compare and contrast different publication formats, and find links to publishing platforms. This tool will assist scientists in getting otherwise inaccessible, hidden data out of the file drawer into the scientific community and literature. We briefly highlight essential details that should be included to ensure reporting quality, which will allow others to use and benefit from research published in these new formats.

The ever-expanding file drawer: where data go to die

Many laboratories have a ‘file drawer’ [1] of unpublished data from well-designed experiments. There are many reasons why data may end up in the file drawer [2,3]. For example, the research team may not have the time or expertise required to analyze the entire data set. Lower priority datasets may remain unpublished, as lab members focus on preparing manuscripts containing the results of high priority experiments. Parts of the study may be missing or incomplete. The study may be a failed replication attempt. Personnel responsible for the project may have left the laboratory before writing a manuscript, or the authors may have published some parts from a larger study, but not others. Alternatively, editors may have rejected the manuscript because the findings were not exciting enough for publication in the authors’ journal of choice.

Regardless of the reasons, failing to publish data from well-designed experiments creates problems for individual researchers, the scientific community and the public. Scientists in preclinical and translational

*These authors contributed equally to this work.

Received: 27 August 2020
Revised: 30 September 2020
Accepted: 02 October 2020

Version of Record published:
28 October 2020

research have invested time and research funds to design and conduct studies yielding valuable data that they have either chosen not to publish, have not been able to publish, or have only partially published [4–6]. Funding agencies and the public do not learn anything from research that is not shared; hence the resources used to complete the work are wasted [7]. Other laboratories, who have no way of knowing that the research was ever conducted, may invest additional time and funding to repeat the same types of studies. Additional problems depend on the type of study. When data from animal studies are not published or shared, animals suffer or are killed without benefits to scientists or society [8–10]. Publication bias can also create risks for patients. Unpublished preclinical data can lead to flawed decisions about whether a potential therapy should advance to clinical trials, exposing patients to unnecessary burdens and risks [11]. When results from neutral or negative clinical trials are not published, clinicians' decisions and recommendations about patient care are based on incomplete evidence [11].

Studies with neutral and null results are more likely to end up in the file drawer than studies with statistically significant findings [12]. This publication bias (Box 1) leaves scientists, funding agencies and clinicians with a distorted view of the scientific evidence, which can lead to poor decisions about what research directions are most promising and should be funded or what medical treatments should be recommended to patients [13]. Such practices can have detrimental consequences. During the 1980s, over 100,000 people died after receiving lorcaïnide-class like drugs. These antiarrhythmic medications were routinely prescribed to patients after a heart attack. A publication with data on the lethal side effects of lorcaïnide was repeatedly rejected and ultimately not published for 13 years, as the authors did not interpret the death rates in their small study as conclusive evidence and journals repeatedly refused to publish these null results [14–16]. While this example is extreme, it illustrates the potential harmful effects of publication bias. Selective reporting of results can create similar problems.

Box 1 Publication bias and selective reporting

- Publication bias occurs when study results influence decisions by authors, reviewers or editors about whether to publish a study, independent of the quality of the research.
- Publication bias distorts scientists' perception of the evidence. When studies showing an effect are more likely to be published than those with null results, a meta-analysis may incorrectly conclude that there is an effect, or may overestimate the effect size [17]. The potential for distortion increases as the probability of publishing null or neutral results decreases.
- Several factors contribute to publication bias, and influence the degree to which publication bias distorts the scientific evidence [13].
 - **Prioritizing statistical significance:** The incorrect belief that statistically significant findings are important and relevant, whereas findings that are not statistically significant are less important and less relevant, contributes to publication bias. These beliefs can affect authors' decisions about whether to submit a manuscript, or editors' and reviewers' decisions to recommend publication of the manuscript.
 - **Prior publications:** Researchers who hypothesized that there was an effect based on published studies may erroneously conclude that their study design, methods or results were faulty if the hypothesized effect is not found and avoid submitting their study for publication. These beliefs emphasize statistical significance and agreement with previous results over effect sizes and study quality.
 - **Effect size:** When the effect is large, most studies will yield statistically significant results. Studies with null or neutral results will be uncommon; hence fewer studies will remain unpublished due to publication bias. When the effect size is small, publication bias is a bigger problem. Many studies will yield negative or neutral results and may be subject to publication bias [2].
 - **Statistical power:** Publication bias may be a greater problem in fields where researchers typically conduct small, underpowered studies [2]. Assuming that there is an effect, high-powered studies are more likely to detect this effect than low powered studies.
- Another related problem is selective reporting. Publication bias occurs when scientists make decisions about whether to publish an entire study based on the results. Selective reporting occurs when authors, reviewers or editors make decisions about whether to publish particular outcome variables

based on the results. Authors may decide, for example, to ‘selectively report’ measurements with statistically significant differences and omit variables that were not statistically different. Alternatively, authors may be more likely to report parts of an experiment that support their hypothesis and less likely to publish parts of an experiment that do not support their hypothesis.

- Selective reporting can also occur when authors provide more information about statistically significant results, compared to non-significant results. For example, authors may report detailed information about statistically significant findings in tables and figures, but state that data were not shown for non-significant findings. When summary statistics and sample sizes are not available, data cannot be replicated or included in a meta-analysis.
- Determinants of selective reporting include a focus on preferred findings, poor or flexible research design, publishing in fields with a high risk of selective reporting, dependence upon sponsors, prejudice, and other factors [3].

Unfortunately, scientists don’t know what proportion of data are never published because there is no comprehensive registry of all planned studies. A study in social science found that two thirds of survey-based experiments that produced null results ended up in the file drawer, whereas nearly all experiments with statistically significant results supporting the underlying hypothesis were published [18]. Many reports confirm the same phenomenon in the medical field, where negative results are less likely to be published [12,19,20]. Clinical trials offer another unique opportunity to assess publication bias, as journal editors began requiring trial registration in 2005 [21,22]. Estimates from AllTrials (<http://www.alltrials.net/>), based on comparisons of registered versus published trials, suggest that approximately 50% of clinical trials results remain unpublished [23,24].

New solutions

New publication formats (Box 2) make it easier for scientists to share research, regardless of the outcome (Table 1), while also ensuring that the data become a part of the permanent scientific record. Tables 1 and 2 and *fiddle* explain and compare these different formats. *fiddle* also provides links to websites for publishers of each publication format, which researchers can use to find sample publications that may be relevant to their field. *Fiddle* focuses on generalist publishers that publish papers from many different fields; it does not provide a comprehensive list of discipline specific publishers. The tool does include links to curated lists designed to help readers identify specialized repositories and discipline-specific databases (i.e. re3data.org, fairsharing.org, and Nature’s list of recommended repositories). Users who are interested in discipline specific repositories or databases can use these links to identify suitable options once they have chosen a publication format.

Box 2 What counts as publication?

In this paper, “publication” refers to any documented product derived from research data that is in the public domain. The various publication formats described in *fiddle* differ in the degree of documentation and intellectual reflection. A traditional journal article accompanied by archived raw data has the highest degree of data set enrichment, whereas data deposited in a repository contains the smallest degree of enrichment. The six publication formats described in *fiddle* have four things in common. They are all: 1. Assigned a permanent digital object identifier (DOI), 2. Findable via different scholarly indexing tools and many provide open access availability 3. Citable and 4. Attributable to an author or originator.

Some of the publication formats in *fiddle* complement one another, or can be used to enhance traditional publications. For instance, datasets deposited in repositories complement traditional research articles published as preprints, on publishing platforms, or in peer reviewed journals. An increasing number of peer-reviewed journals simplify the submission process by allowing authors to directly submit preprints to the journal for consideration.

As the use of preprints and other alternative publication formats continues to grow, the incentives for avoiding publication bias and using new publication formats will continue to shift [25]. The Declaration of Helsinki notes that all researchers have an ethical obligation to disseminate research results

Table 1 *fiddle* allows authors to quickly compare and contrast different publication formats

	Data Repository	Micropublication	Preprint publication	Data journals	Publishing platform	Journal open to null results
Description	Platforms that allow upload of research datasets to make them citable and reusable.	Designed for unpublished observations, negative/neutral results that do not require a scientific narrative.	Platforms for unpublished research manuscripts that allow others to immediately view the manuscript.	Journal article that focuses on presenting a dataset with metadata and the methods used to acquire the dataset.	Articles are published without editorial filtering; peer-review happens after (immediate) publication of the article.	Traditional journals that also publish null results.
Providers	Zenodo, FigShare or Dryad; to search for disciplinary repositories use re3data, fairsharing, or Nature's list	ScienceMatters, BMC Research Notes	bioRxiv, medRxiv, osf.io	Scientific Data, Data, Data in Brief, F1000 Data Note, many disciplinary journals (e.g. GigaScience)	F1000Research, Open Research Central	PeerJ, PLoS One, Scientific Reports, multiple BMC journals and many other disciplinary journals
Effort	low effort	low effort	medium effort	some effort to prepare manuscript/data	some effort to prepare manuscript/data	some effort to prepare manuscript/data
Costs in EUR	free of charge	600 - 1300 €	free of charge	up to 1500 €	up to 1000 €	up to 1600 €
Costs in US\$	free of charge	670 - 1440 \$	free of charge	up to 1670 \$	up to 1100 \$	up to 1780 \$
Time to publication	immediate	typically 1-3 months	immediate	typically 1-4 months	immediate	typically 1-6 months
Recognition	citations of the dataset	citations of article, article can be listed in CV (future handling of such articles is open)	citations of article, article can be listed in CV (not universally accepted at this point)	citations of article, article can be listed in CV	citations of article, article can be listed in CV (not universally accepted at this point)	citations of article, article can be listed in CV
Publishing venue can have Impact Factor	no	yes	no	yes	no	yes
Peer-review	no	peer-review	post-publication review possible	peer-review	peer-review	peer-review
DOI	yes	yes	yes	yes	yes	yes
Versioning	yes	no	yes	yes	yes	no
Indexing:						
Pubmed	no	no	no	yes	yes	yes
Pubmed Central	no	some	no	some	yes	Yes
Web of Science	no	some	no	most	no	yes
Scopus	no	some	no	some	no	yes
CrossRef	no	some	yes	some	yes	yes
Google Scholar	no	yes	yes	yes	yes	yes
Additional information		integrated open data upload, reviewer compensation, often only one reviewer	preprint deposit accepted by large majority of journals and often offered as integral steps in submission process (see Sherpa/ROMeO)			

The first two rows of *fiddle* describe each publication format and offer links to providers or publishers. The remaining columns allow users to compare publication formats according to different characteristics (required effort, cost, whether materials are peer reviewed, what databases index materials, etc.)

Table 2 What do different publication formats include?

Information	Data Repository	Micropublication	Preprint	Data Journal	Publishing platform	Journal Open to Null Results
Abstract	✓	✓	✓	✓	✓	✓
Introduction		Brief	✓	✓	✓	✓
Methods		Brief	✓	✓	✓	✓
Results		Brief	✓		✓	✓
Discussion and interpretation			✓		✓	✓
Raw data	✓			✓		
Metadata	✓			✓		
Peer reviewed	No	Yes	No*	Yes	Yes‡	Yes

The table provides a rough overview of what different publishing formats include, as well as information on whether the format is typically peer reviewed. Check marks indicate that the publication format traditionally includes the item, whereas blank spaces indicates that the publication format does not traditionally include this item. ‘Brief’ indicates that the publication includes a condensed version of this item. See Table 1 and *fiddle* for additional information on each format.

*Post-publication peer review of pre-prints is possible

‡Peer review for publishing platforms happens after immediate posting of the article

These article types do not traditionally include raw data or metadata, however raw data and metadata can be deposited in a data repository and cited in preprints, or papers posted on publishing platforms and in journals open to null results.

[26]. Funders have highlighted the importance of ensuring that research outputs, including negative results, are published [7,27,28]. Funding agencies such as the National Institutes of Health in the United States allow researchers to cite preprints in grant applications [29]. Papers that deposit open data accumulate up to 25% more citations than papers that do not have open data [30,31].

Some of the new publication formats follow the same format as traditional peer-reviewed research articles, but make it easier to publish manuscripts that would typically be rejected from journals where editorial and peer review often prioritizes exciting results. Preprints, for example, are unpublished manuscripts that have not been peer-reviewed and are shared immediately with the scientific community. Platforms that publish preprints include bioRxiv, medRxiv, and Open Science Framework Preprint Services. Publication platforms, such as F1000 Research and Open Research Central, publish articles immediately without editorial filtering. Open peer review occurs after publication. Journals that are open to null results are traditional journals that publish peer-reviewed manuscripts but welcome all studies regardless of outcome. Such journals have clear public policies to publish manuscripts describing well-designed studies with null results, results that appear to contradict those of previous publications [32] or other research outcomes that are hard to publish. Examples include PeerJ and PLOS ONE.

Other new publication formats facilitate publication of data or results that would be difficult to share in an Introduction-Method-Results-Discussion format. Micropublications, for example, are very short publications designed for unpublished observations, neutral or null results or other research that does not require a scientific narrative [33]. Platforms that publish micropublications include Science Matters and BMC Research Notes. Data repositories, such as figshare, Zenodo or Dryad, allow scientists to upload small or large research datasets to make them citable and reusable. Data journals, such as Scientific Data or Data, publish journal articles that present a dataset, metadata explaining the dataset and the data collection methods.

As the use of preprints [25] and other alternative publication formats continues to grow, even researchers who choose not to use these formats will benefit from understanding how they work. Knowing where these research outputs are indexed and whether they are peer reviewed, for example, is essential to finding and evaluating materials that are relevant to one’s area of research. Scientists who are unaware of data repositories and data journals may miss opportunities to use datasets relevant to their work. Researchers who don’t know about rapidly growing preprint servers may not find out about important studies until papers are published, often many months after the preprints were first posted. Tools that help researchers to understand different publication formats and identify those formats that are most appropriate for the dissemination of their data are thus urgently needed [34].

Options

Choose the criteria that are most important to you.

I have a

unanalyzed dataset

stand-alone finding or results from a small study

results from a full-scale study

rejected manuscript

Money I have to cover publication cost

0

up to 1000 EUR/1100 US\$

up to 2000 EUR/2200 US\$

I want the outcome to be indexed in

Pubmed

Pubmed Central

Web of Science

Scopus

CrossRef

Google scholar

I want my outcome to be peer-reviewed

any

yes

no

I want my publication or dataset to appear immediately

yes

no

If these options aren't helpful, try our list of scenarios instead:

[Go to scenarios](#)

Scenarios

Why are your data in the file drawer?

I don't have enough time to prepare a publication

My experiment or dataset is incomplete

I have data that may be useful to others, but don't have time to analyze everything

I have neutral or null results from a small, underpowered study or an exploratory study

I have neutral or null results from a large, adequately powered study

My study is completed, but the findings aren't novel or exciting

I need the research to be published quickly

I don't have funding to pay for publication charges

None of these describe my situation - show me the table of all options

[Back to options](#)

Figure 1. Search strategies in *fiddle*

Authors can identify publication formats that meet their needs by selecting either characteristics that are most important or relevant to them (Search by Options) or by selecting the scenario that best describes their situation (Search by Scenarios).

fiddle: the file drawer data liberation effort tool

fiddle is a free, open source 'matchmaking' tool designed to help researchers to identify the publication format that will work best for a particular dataset or study that may be hard to publish in traditional journals (RRID:SCR_017327, available at: <http://s-quest.bihealth.org/fiddle/>). The tool includes a link to a brief video tutorial. Researchers can use this shiny (RRID:SCR_001626) [35] app to quickly compare characteristics of different publications formats and search for a format that best meets their needs. *Fiddle* is not discipline-specific and can be used for any life science field where publication occurs and where research results from well designed and executed studies remain hidden in the file drawer. Once users have identified a publication format in *fiddle*, they can click on links to visit websites of relevant publishers or platforms, or see examples of this particular format.

There are two ways to search for publication formats (Figure 1). The first filtering option is to search by important characteristics describing the dataset and the researcher's publishing-related preferences. Users can find suitable publishing platforms by answering the questions below that are most relevant to them:

1. What type of unpublished information do you have (unanalyzed dataset, rejected manuscript, etc.)?
2. Amount of funding available for publication costs

Table 3 Guidelines for conducting transparent, rigorous and reproducible research

Guidelines for common types of studies			
Study type	Guideline acronym	RRID or Citation	Link
Observational studies	STROBE	RRID: SCR.018788	https://www.strobe-statement.org/
Animal studies - planning	PREPARE	RRID:SCR.018787	https://norecopa.no/PREPARE
Animal studies - reporting	ARRIVE 2.0	RRID:SCR.018719	https://arriveguidelines.org/arrive-guidelines
Randomized controlled trials	CONSORT	RRID:SCR.018720	http://www.consort-statement.org/
Systematic review and meta-analysis	PRISMA	RRID:SCR.018721	http://www.prisma-statement.org/
Systematic review and meta-analysis of observational studies	MOOSE	Stroup et al., 2000 [36]	https://jamanetwork.com/journals/jama/fullarticle/192614
Consult these resources to find guidelines for other types of studies			
Resource description	Resource name	RRID	Link
Guidelines for many different types of studies	EQUATOR network	RRID:SCR.012861	https://www.equator-network.org/

The table provides information on guidelines for specific types of studies that are common in many fields, as well as resources that will allow researchers to find guidelines for less common types of studies.

- Where should the publication or dataset be indexed?
- Do you want the publication or dataset to be peer reviewed?
- Do you want the publication or dataset to appear immediately?

The other, alternative filtering option is to search by scenarios that describe the reason why the information is unpublished. Example scenarios include ‘I don’t have enough time to prepare a publication,’ ‘I have data that may be useful to others, but am not able to analyze everything,’ and ‘My study is completed, but the findings aren’t novel or exciting’. The tool highlights publication formats that meet the user’s requirements. Users can review detailed information on each type of publication format and then click on links to visit websites for different publishers. Users can also compare all publication formats. All options in *fiddle* provide a permanent, citable and findable link to the data or manuscript. Many formats are also peer-reviewed. The source code for *fiddle* is available at <https://github.com/quest-bih/fiddle>.

Should all data from the file drawer be published?

Scientists should focus on the quality of the study methods, rather than the desirability of the results, when deciding which file drawer data to publish. While data from well-designed experiments often ends up in the file drawer, many file drawers also contain data from poorly designed, badly conducted or insufficiently documented experiments that are unlikely to be reproducible or useful. *fiddle* and other efforts to reduce publication bias and selective reporting encourage authors to publish data from well-designed experiments that may be useful to the scientific community or the public, regardless of whether the findings were statistically significant. This applies also for datasets that are too small to yield reliable conclusions, however may be informative when combined with many other datasets using techniques such as meta-analysis.

fiddle is not intended to promote publication of data from poor quality studies that are unlikely to be useful or informative. Authors should consult study design and reporting guidelines when designing studies and preparing publications to increase the likelihood that data will be transparent, rigorous and reproducible. Table 3 lists guidelines for common types of studies in many fields, including observational studies, animal studies, randomized controlled trials, and systematic reviews and meta-analyses. Guidelines for other types of studies can be found through the EQUATOR network website (RRID:SCR.012861).

What is needed to ensure that the data are useful to others?

The goal of publishing file drawer data is to make these research outputs available to the scientific community; therefore, scientists should ensure that the information is shared in a form that others can understand and use. The list below outlines some important features that should be reported for most, if not all, formats listed in *fiddle*. Additional information may be needed, depending on the publication format, study design, experimental methods, and type of

data that is generated. The lack of time is one reason for not publishing file drawer data [3]; therefore there may be trade-offs between efforts to reduce publication bias by introducing shorter publication formats that take less time to prepare and attempts to improve transparency and reproducibility by encouraging authors to report detailed information required to assess study quality. Information that scientists need to interpret and use scientific data include the following:

1. **Research question:** The material provided should clearly specify the research question that the study was designed to answer, along with any hypotheses.
2. **Participants, subjects, specimens or samples:** The material should specify who the participants or subjects were, and how specimens or samples were obtained. When appropriate, the authority that gave regulatory study approval should be stated (i.e. institutional review board, animal care and use committee, etc.). Human studies should state how patients were consented.
3. **Study design:** The material should specify the study design, and state whether the study was exploratory or confirmatory. Important design features needed to assess the risk of bias should be reported. These include whether the measurements and analyses were performed in a blinded fashion, whether participants or subjects were randomized to the different conditions and how randomization was performed, a power calculation or sample size justification, and details on the number of excluded observations and reasons for exclusion [37,38].
4. **Data:** A scientist without prior knowledge of the experiment should be able to interpret and use the dataset based on the meta-data provided. The dataset should be compliant with the respective Minimum Information for Biological and Biomedical Investigations (RRID:SCR_002042, <https://fairsharing.org/collection/MIBBI>) and include a data dictionary that clearly explains what each variable is, what the measurement units are and how the variables were measured (<https://dataedo.com/blog/different-types-of-tools-you-can-use-to-create-data-dictionary>). Data should have a license specifying any conditions for re-use. Authors who share data should consult the FAIR data principles [39] and plan their data documentation [40]. When depositing data obtained from human samples or patient data, regional data protection laws and legislations apply and need to be considered prior to the start of the project to found out which form of consent, de-identification procedures or data access restrictions may apply. Some research institutions employ a data protection or open data specialist to help researchers with open data issues. Investigators working with patient data should contact their institutional review board for guidance.
5. **Results:** Readers should know what was measured, be able to determine sample sizes for each group and/or analysis and know what summary statistics are reported.
6. **Analysis:** If the data were analyzed, the material should provide enough information to determine how the analysis was conducted. This could include code for the analysis. The SAMPL guidelines [41] recommend providing enough detail so that a reader who understands statistics could reproduce the analysis if he or she had access to the data.
7. **Limitations:** The limitations of the data or study should be clearly explained.
8. **Contact person:** If the uploading author is not the best person to answer additional questions, the name and contact information for one or two people with such knowledge should be provided.

What can scientists do to prevent distortion of the scientific literature due to publication bias and selective reporting?

Researchers can take several steps to reduce publication bias and accelerate scientific discovery. The first step is to plan ahead. Research teams should ask all collaborators to commit to publishing all study results, regardless of the perceived importance of the results and whether the results support the hypothesis. An additional strategy is to pre-register a study by posting a publicly available, time stamped protocol that outlines the study objectives and hypotheses, data collection procedures and planned analyses. Cite this pre-registration when publishing the study, regardless of which publication format is used, and provide an explanation if the final study differs from the pre-registered protocol. Pre-registration addresses publication bias by allowing researchers to identify studies that were conducted, but not published. Studies can be pre-registered on sites like AsPredicted (RRID:SCR_018789, <https://aspredicted.org>) and the Open Science Framework (RRID:SCR_003238, <https://osf.io>)

Once the study is complete, researchers should share all findings with the scientific community using traditional peer-reviewed publications or alternative publication formats described in *fiddle*. Specific actions that researchers can take include using repositories and other platforms to share data and protocols, and avoiding 'data not shown' statements. Scientists who have a sound scientific reason not to report data should specify this when publishing or sharing such study results. Investigators might report, for example, that one variable measured was not reported due to device malfunctioning on the day the test was performed. Finally, scientists should talk to their colleagues about the consequences of publication bias and selective reporting. These conversations are especially important when co-authors, reviewers or editors encourage selective reporting.

Conclusions

The open source *fiddle* tool is a match-making Shiny app designed to help researchers identify the publication format that is most appropriate for their publication or dataset. Users can search for a publication format that meets their needs, compare and contrast different publication formats, and find links to publishers and examples. This tool will assist scientists in getting otherwise inaccessible data from well-designed experiments out of the file drawer and into the scientific community to reduce bias in the scientific literature. Finally, funding agencies, journals, and hiring and promotion committees need to incentivize and reward publication of all research from well-designed experiments, regardless of the form of publication. Some investigators may be reluctant to publish studies that are unlikely to be accepted by journals with high impact factors due to concerns that funding agencies or promotion and tenure committees may devalue this work, adversely affecting career advancement. This perception bolsters publication bias by encouraging scientists to publish only their most interesting and impactful research, to the detriment of the scientific community and the public. We hope that this paper and the tool will raise awareness of the negative consequences of publication bias and selective reporting, and encourage the scientific community to work towards individual and systemic change.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

R.B. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2049 – 390688087. T.L.W. was funded by American Heart Association [grant number 16GRNT30950002]. This publication was made possible by CTSA [grant number UL1 TR000135] from the National Center for Advancing Translational Sciences, a component of the National Institutes of Health. The content is solely the authors' responsibility and does not necessarily represent the official views of the NIH. The writing of the manuscript and the decision to submit it for publication were solely the authors' responsibilities.

Abbreviation

fiddle, file drawer data liberation effort.

References

- 1 Rosenthal, R. (1979) The file drawer problem and tolerance for null results. *Psychol. Bull.* **86**, 638–641, <https://doi.org/10.1037/0033-2909.86.3.638>
- 2 Song, F. et al. (2010) Dissemination and publication of research findings: an updated review of related biases. *Health Technol. Assess. (Rockv.)* **14**, 1–193, <https://doi.org/10.3310/hta14080>
- 3 van der Steen, J.T. et al. (2018) Determinants of selective reporting: A taxonomy based on content analysis of a random selection of the literature. *PLoS ONE* **13**, e0188247, <https://doi.org/10.1371/journal.pone.0188247>
- 4 Chiu, K., Grundy, Q. and Bero, L. (2017) Spin' in published biomedical literature: A methodological systematic review. *PLoS Biol.* **15**, e2002173, <https://doi.org/10.1371/journal.pbio.2002173>
- 5 Matosin, N. et al. (2014) Negativity towards negative results: a discussion of the disconnect between scientific worth and scientific culture. *Dis. Model Mech* **7**, 171–3, <https://doi.org/10.1242/dmm.015123>
- 6 McElreath, R. and Smaldino, P.E. (2015) Replication, Communication, and the Population Dynamics of Scientific Discovery. *PLoS ONE* **10**, e0136088, <https://doi.org/10.1371/journal.pone.0136088>
- 7 Collins, E. (2013) Publishing priorities of biomedical research funders. *BMJ Open* **3**, e004171, <https://doi.org/10.1136/bmjopen-2013-004171>
- 8 Conradi, U. and Joffe, A.R. (2017) Publication bias in animal research presented at the 2008 Society of Critical Care Medicine Conference. *BMC Res. Notes* **10**, 262, <https://doi.org/10.1186/s13104-017-2574-0>
- 9 Strech, D. and Dirnagl, U. (2019) 3Rs missing: animal research without scientific value is unethical. *BMJ Open Sci.* **3**, bmjos-2018-000048, <https://doi.org/10.1136/bmjopen-2018-000048>

- 10 Wieschowski, S. et al. (2019) Publication rates in animal research. Extent and characteristics of published and non-published animal studies followed up at two German university medical centres. *PLoS ONE* **14**, e0223758, <https://doi.org/10.1371/journal.pone.0223758>
- 11 Yarborough, M. et al. (2018) The bench is closer to the bedside than we think: Uncovering the ethical ties between preclinical researchers in translational neuroscience and patients in clinical trials. *PLoS Biol.* **16**, e2006343, <https://doi.org/10.1371/journal.pbio.2006343>
- 12 Fanelli, D. (2012) Negative Results Are Disappearing from Most Disciplines and Countries. *Scientometrics* **90**, 891–904, <https://doi.org/10.1007/s11192-011-0494-7>
- 13 Mlinaric, A., Horvat, M. and Supak Smolic, V. (2017) Dealing with the positive publication bias: Why you should really publish your negative results. *Biochem Med. (Zagreb)* **27**, 030201, <https://doi.org/10.11613/BM.2017.030201>
- 14 Bruckner, T. and Ellis, B. (2017) Clinical Trial Transparency: A Key to Better and Safer Medicines. [cited 2019 06-27]; Available from: https://media.wix.com/ugd/01f35d_0f2955eb88e34c02b82d886c528efeb4.pdf
- 15 Dickersin, K. and Rennie, D. (2003) Registering clinical trials. *JAMA* **290**, 516–23, <https://doi.org/10.1001/jama.290.4.516>
- 16 Moor, T. (1995) *Deadly Medicine: Why Tens of Thousands of Heart Patients Died in America's Worst Drug Disaster*, 1st edn., p. 352, Simon & Schuster, New York
- 17 Michel, M.C., Murphy, T.J. and Motulsky, H.J. (2020) New Author Guidelines for Displaying Data and Reporting Data Analysis and Statistical Methods in Experimental Biology. *Drug Metab. Dispos.* **48**, 64–74, <https://doi.org/10.1124/dmd.119.090027>
- 18 Franco, A., Malhotra, N. and Simonovits, G. (2014) Publication bias in the social sciences: Unlocking the file drawer. *Science* **345**, 1502–1505, <https://doi.org/10.1126/science.1255484>
- 19 Duyx, B. et al. (2019) The strong focus on positive results in abstracts may cause bias in systematic reviews: a case study on abstract reporting bias. *Syst Rev* **8**, 174, <https://doi.org/10.1186/s13643-019-1082-9>
- 20 Jannot, A.S. et al. (2013) Citation bias favoring statistically significant studies was present in medical research. *J. Clin. Epidemiol.* **66**, 296–301, <https://doi.org/10.1016/j.jclinepi.2012.09.015>
- 21 (2004) Clinical trial registration: a statement from the International Committee of Medical Journal Editors. [cited 2019 April 10]; Available from: http://www.icmje.org/news-and-editorials/clin_trial_sep2004.pdf
- 22 Laine, C. et al. (2007) Update on Trials Registration: Clinical Trial Registration: Looking Back and Moving Ahead. [cited 2019 April 10]; Available from: http://www.icmje.org/news-and-editorials/clinical_trial_reg_jun2007.html#ref
- 23 AllTrials. How many clinical trials are left unpublished?, [cited 2019 April 10]; Available from: <http://www.alltrials.net/wp-content/uploads/2016/03/AllTrials-briefing-doc-how-many-clinical-trials-unreported.pdf>
- 24 Sena, E.S. et al. (2010) Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol.* **8**, e1000344, <https://doi.org/10.1371/journal.pbio.1000344>
- 25 Abdill, R.J. and Blekhman, R. (2019) Tracking the popularity and outcomes of all bioRxiv preprints. *Elife* **8**, e45133, <https://doi.org/10.7554/eLife.45133>
- 26 World Medical Association. (2013) World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* **310**, 2191–4, <https://doi.org/10.1001/jama.2013.281053>
- 27 Galsworthy, M.J. et al. (2012) Academic output of 9 years of EU investment into health research. *Lancet* **380**, 971–2, [https://doi.org/10.1016/S0140-6736\(12\)61528-1](https://doi.org/10.1016/S0140-6736(12)61528-1)
- 28 Riley, W.T., Riddle, M. and Lauer, M. (2018) NIH Policies on Experimental Studies with Humans. *Nat. Hum. Behav.* **2**, 103–106, <https://doi.org/10.1038/s41562-017-0265-4>
- 29 Kaiser, J. (2017) NIH enables investigators to include draft preprints in grant proposals. *Science* n.pag, <https://doi.org/10.1126/SCIENCE.AAL0961>
- 30 Piwowar, H.A. and Vision, T.J. (2013) Data reuse and the open data citation advantage. *PeerJ* **1**, e175, <https://doi.org/10.7717/peerj.175>
- 31 Colavizza, G. et al. (2019) The citation advantage of linking publications to research data. [cited 2019 12.31.2019]; Available from: <https://arxiv.org/abs/1907.02565v2>
- 32 Kannan, S. and Gowri, S. (2014) Contradicting/negative results in clinical research: Why (do we get these)? Why not (get these published)? Where (to publish)? *Perspect. Clin. Res.* **5**, 151–3, <https://doi.org/10.4103/2229-3485.140546>
- 33 Raciti, D., Yook, K., Harris, TW, Schedl, T and Sternberg, PW (2018) Micropublication: incentivizing community curation and placing unpublished data into the public domain. *Database (Oxford)* bay013, <https://doi.org/10.1093/database/bay013>
- 34 Iwema, C.L. et al. (2016) search.bioPreprint: a discovery tool for cutting edge, preprint biomedical research articles. *F1000Res* **5**, 1396, <https://doi.org/10.12688/f1000research.8798.1>
- 35 Chang, W. et al. (2018) shiny: Web Application Framework for R. R package version 1.1.0.
- 36 Stroup, D.F. et al. (2000) Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting. *JAMA* **283**, 2008–2012, <https://doi.org/10.1001/jama.283.15.2008>
- 37 (2019) Reviewer Guidance on Rigor and Transparency: Research Project Grant and Mentored Career Development Applications. *NIH Peer Rev.*, [PDF] 03-18-19 [cited 2019 06-18-19]; Available from: https://grants.nih.gov/grants/peer/guidelines_general/Reviewer_Guidance_on_Rigor_and_Transparency.pdf
- 38 Landis, S.C. et al. (2012) A call for transparent reporting to optimize the predictive value of preclinical research. *Nature* **490**, 187–91, <https://doi.org/10.1038/nature11556>
- 39 Wilkinson, M.D. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018, <https://doi.org/10.1038/sdata.2016.18>
- 40 Hart, E.M. et al. (2016) Ten Simple Rules for Digital Data Storage. *PLoS Comput. Biol.* **12**, e1005097, <https://doi.org/10.1371/journal.pcbi.1005097>

41 Lang, T. and Altman, D. (2013) Basic statistical reporting for articles published in clinical medical journals: the Statistical Analyses and Methods in the Published Literature, or SAMPL guidelines. [cited 2018 December 20]; Available from:
<http://www.equator-network.org/wp-content/uploads/2013/07/SAMPL-Guidelines-6-27-13.pdf>