




## CASE STUDY

# Poor statistical reporting, inadequate data presentation and spin persist despite Journal awareness and updated *Information for Authors* [version 1; peer review: 2 approved, 2 approved with reservations]

Martin Héroux <sup>1,2</sup>, Joanna Diong<sup>1,3</sup>, Elizabeth Bye<sup>1,2</sup>, Georgia Fisher<sup>1,4</sup>, Lucy Robertson<sup>1,5</sup>, Annie Butler<sup>1,2</sup>, Simon Gandevia<sup>1,5</sup>

<sup>1</sup>School of Biomedical Sciences, University of New South Wales, Sydney, New South Wales, 2052, Australia

<sup>2</sup>Neuroscience Research Australia, Sydney, NSW, 2031, Australia

<sup>3</sup>School of Biomedical Sciences, The University of Sydney, Sydney, New South Wales, 2006, Australia

<sup>4</sup>Faculty of Medicine, Health and Human Sciences, Macquarie University, Sydney, New South Wales, 2109, Australia

<sup>5</sup>School of Clinical Medicine, University of New South Wales, Sydney, New South Wales, 2031, Australia

**V1** First published: 20 Nov 2023, 12:1483  
<https://doi.org/10.12688/f1000research.142841.1>





Latest published: 20 Nov 2023, 12:1483  
<https://doi.org/10.12688/f1000research.142841.1>



## Abstract

Sound reporting of research results is fundamental to good science. Unfortunately, poor reporting is common and does not improve with editorial educational strategies. We investigated whether publicly highlighting poor reporting at a journal can lead to improved reporting practices. We also investigated whether reporting practices that are required or strongly encouraged in journal *Information for Authors* are enforced by journal editors and staff. A 2016 audit highlighted poor reporting practices in the Journal of Neurophysiology. In August 2016 and 2018, the American Physiological Society updated the *Information for Authors*, which included the introduction of several required or strongly encouraged reporting practices. We audited Journal of Neurophysiology papers published in 2019 and 2020 (downloaded through the library of the University of New South Wales) on reporting items selected from the 2016 audit, the newly introduced reporting practices, and items from previous audits. Summary statistics (means, counts) were used to summarize audit results. In total, 580 papers were audited. Compared to results from the 2016 audit, several reporting practices remained unchanged or worsened. For example, 60% of papers erroneously reported standard errors of the mean, 23% of papers included undefined measures of variability, 40% of papers failed to define a

## Open Peer Review

Approval Status 

	1	2	3	4
<b>version 1</b>				
20 Nov 2023	<a href="#">view</a>	<a href="#">view</a>	<a href="#">view</a>	<a href="#">view</a>

1. **Ulf Kahlert** , Medical Faculty University Hospital, Magdeburg,, Germany
2. **Daniel J. Dunleavy** , Florida State University, Tallahassee, USA
3. **Clarissa F. D. Carneiro** , QUEST Center for Responsible Research,, Berlin, Germany
4. **Christopher McCrum**, Maastricht University Medical Centre, Maastricht, The Netherlands

Any reports and responses or comments on the article can be found at the end of the article.

statistical threshold for their tests, and when present, 64% of papers with p-values between 0.05 and 0.1 misinterpreted them as statistical trends. As for the newly introduced reporting practices, required practices were consistently adhered to by 34 to 37% of papers, while strongly encouraged practices were consistently adhered to by 9 to 26% of papers. Adherence to the other audited reporting practices was comparable to our previous audits. Publicly highlighting poor reporting practices did little to improve research reporting. Similarly, requiring or strongly encouraging reporting practices was only partly effective. Although the present audit focused on a single journal, this is likely not an isolated case. Stronger, more strategic measures are required to improve poor research reporting.

### Keywords

Meta-research, research quality, scientific reporting, reproducibility



This article is included in the [Research on Research, Policy & Culture](#) gateway.

**Corresponding author:** Martin Héroux ([m.heroux@neura.edu.au](mailto:m.heroux@neura.edu.au))

**Author roles:** **Héroux M:** Conceptualization, Data Curation, Formal Analysis, Investigation, Project Administration, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Diong J:** Formal Analysis, Investigation, Software, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Bye E:** Investigation, Validation, Writing – Review & Editing; **Fisher G:** Investigation, Validation, Writing – Review & Editing; **Robertson L:** Investigation, Validation, Writing – Review & Editing; **Butler A:** Formal Analysis, Investigation, Methodology, Validation, Writing – Review & Editing; **Gandevia S:** Investigation, Validation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** The author(s) declared that no grants were involved in supporting this work.

**Copyright:** © 2023 Héroux M *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Héroux M, Diong J, Bye E *et al.* **Poor statistical reporting, inadequate data presentation and spin persist despite journal awareness and updated *Information for Authors* [version 1; peer review: 2 approved, 2 approved with reservations]** F1000Research 2023, 12:1483 <https://doi.org/10.12688/f1000research.142841.1>

**First published:** 20 Nov 2023, 12:1483 <https://doi.org/10.12688/f1000research.142841.1>

## Introduction

Scientific discovery and the translation of findings into practice requires complete, transparent and unbiased reporting. Although fundamental to the scientific method, sound reporting practices are not always adhered to. Despite clear recommendations,<sup>1–11</sup> poor statistical reporting and biased interpretation of results are common.<sup>12–23</sup>

In 2011, the Journal of Physiology and the British Journal of Pharmacology jointly published a series of editorials to educate their readership about statistics and scientific reporting.<sup>24</sup> These editorials covered topics such as statistical terminology and procedures,<sup>25,26</sup> sample size and statistical power,<sup>27,28</sup> and results presentation and interpretation.<sup>29–31</sup> Together, these editorials became reporting guidelines referenced in the *Information for Authors* of these journals. Unfortunately, an audit of papers published before and after these guidelines were implemented revealed no improvement in reporting practices.<sup>32</sup> For example, over 80% of audited papers inappropriately used the standard error of the mean (SEM) to summarise data variability, while 60% of papers with p-values between 0.05-0.1 misinterpreted these as statistical trends (i.e. spin: reporting results so that they are interpreted in a more favourable light). Thus, educational editorials and encouragement from journals are not sufficient to improve reporting practices.

In 2016, a Letter to the Editor called attention to the high prevalence of poor statistical reporting practices in the Journal of Neurophysiology.<sup>33</sup> This audit evaluated all papers published in 2015, and found that 65% of papers inappropriately used the SEM, only 58% of papers reported exact p-values (e.g.  $p=0.021$  versus  $p<0.05$ ), and 57% of papers with p-values between 0.05-0.1 resorted to spin.

To enhance the quality of scientific reporting in its journals, the American Physiological Society revised its *Information for Authors* in August 2016 and June 2018. Several reporting practices were now required or strongly encouraged.

These events presented a unique opportunity. Specifically, we wondered whether publicly highlighting poor reporting practices would motivate a journal to crack down on poor reporting practices. Moreover, we wondered whether reporting practices that were required or strongly encouraged – not merely encouraged as in the case of the Journal of Physiology and the British Journal of Pharmacology – would result in improved reporting practices.

The present study addressed two research questions: 1) Do reporting practices improve when a journal is informed of poor reporting practices? 2) Do publications adhere to reporting practices that are either required or strongly encouraged? We audited papers published in the Journal of Neurophysiology in 2019 and 2020. Audit items included the three items from the original Journal of Neurophysiology audit,<sup>33</sup> two required and two strongly encouraged reporting practices from the updated American Physiological Society Information for Authors, and four items from other audits we have conducted.<sup>32,34</sup>

## Methods

### Eligibility criteria

Original research papers published in the Journal of Neurophysiology in the years 2019 and 2020 were eligible for inclusion. Papers were excluded if they were editorials, reviews, errata, comments or rebuttals. Full-text PDF files of eligible papers were downloaded through the library of the University of New South Wales in January 2022.

### Audit items

#### Items from the original audit

We assessed whether one or more of the following measures were used to summarise the variability of data or the size of an effect: standard error of the mean (SEM), standard deviation (SD), 95% confidence interval (95% CI) or interquartile range (IQR). We also assessed whether any summary statistic was not defined; this typically happens when the  $\pm$  symbol or errors bars on plots are used but are not defined. The *Guidelines for reporting statistics in journals published by the American Physiological Society*,<sup>1</sup> a document that is referenced in the Journal of Neurophysiology *Information for Authors*, states that the SEM should not be used to summarise the variability of data or the size of an effect.

Additionally, we assessed whether papers reported exact p-values consistently, inconsistently, or not at all. The use of exact p-values is recommended by the *Guidelines for reporting statistics in journals published by the American Physiological Society*.<sup>1</sup>

Finally, we assessed whether authors interpreted p-values close to, but above the selected threshold for statistical significance, typically  $p=0.05$ , as statistically significant or statistical trends (i.e. spin). This practice is misleading: more often than not, if additional data are collected, these p-values increase.<sup>3</sup> To be able to compare our results to those of a

more recent audit,<sup>34</sup> we also determined the prevalence of spin across all papers, regardless of whether or not such a p-value was present.

### Items required or strongly encouraged in updated *Information for Authors*

Although several new reporting practices were introduced in the 2016 and 2018 updates to the *Information for Authors* for American Physiological Society journals (underlying data: *Extract from Information for Authors*), we selected four items that were broadly relevant, easily assessed, and, in two cases, related to items from previous audits.

We audited two required reporting practices introduced in August 2016. First, we assessed whether the number of samples or animals that contributed to plotted results were specified in figures or figure legends. Second, if statistical results were reported in a figure or figure legend, we assessed whether the statistical test used was specified in the figure legend.

We also audited two strongly encouraged reporting practices introduced in August 2018. First, we assessed whether data underlying the results were made available upon request or in a public repository. Second, we assessed whether graphical data “showed the range of data points in relation to the mean value and significance (i.e. dot-whisker plots) rather than formats that may mask variability (i.e. bar graphs)” and was “formatted to best convey the variability in the results” (underlying data: *Extract from Information for Authors*). It is unfortunate that the *Information for Authors* specifically reference ‘dot-whisker plots’ given that a mean (dot) and a two-sided error bar (whisker) is not different from a bar graph. That is, a dot-whisker plot can mask variability just as much as a bar graph. Nevertheless, because this reporting practice was introduced to encourage authors to convey the variability of plotted data and results, we scored this item as ‘yes’ when figures met this requirement: if the individual data points used to generate the summary statistics were plotted (regardless of the type of summary statistic used), or if box-and-whisker (median, IQR, range) or violin plots were encouraged.

### Items from previous audits

To broaden the scope of the present audit and to allow for direct comparison with our previous results, we selected four additional items from two previous audits.<sup>32,34</sup>

We assessed whether a sample size calculation was performed prior to the start of the study. In confirmatory studies, it is important to determine *a priori* the sample size required to achieve sufficient statistical power or generate precise estimates of the investigated effects.<sup>4,35,36</sup>

We also assessed whether the study protocol was registered prior to the start of the study. Study registration is one of the simplest ways to improve the rigour of a study; it combats the selective reporting of outcomes, the cherry-picking of results, and p-hacking.<sup>37–41</sup>

Next, we assessed whether plots included individual data points. As indicated in the *Journal of Neurophysiology Information for Authors*, “the same bar graph can represent many different datasets” (underlying data: *Extract from Information for Authors*). Similarly, in their series of educational editorials published in the *Journal of Physiology* and the *British Journal of Pharmacology*, Drummond & Vowler<sup>30</sup> encourage researchers to “use methods of data presentation that allow inspection, not concealment, of the nature of the distribution of the data.” While closely related to the audit item on variability of plotted data and results, the present item was included to allow for a direct comparison with previous audit results.

Finally, we assessed if a probability threshold was defined when frequentist statistics were used. This recommendation comes from the *Guidelines for reporting statistics in journals published by the American Physiological Society*<sup>1</sup>: “Define and justify a critical significance level appropriate to the goals of your study”.

### Scoring manual and pilot testing

A scoring manual was created to increase the consistency across investigators (extended data: *Scoring manual*). Next, five eligible papers were selected randomly and audited by all investigators. A consensus meeting was held and agreement across investigators was assessed. Wording of audit items and the scoring manual was refined to improve the consistency and accuracy of audit results.

### Data extraction and analysis

Each investigator was randomly allocated ~80 papers to audit, which they assessed independently.

If investigators were unsure how to score an audit item for a given paper, it was discussed amongst the team and agreement was reached by consensus.

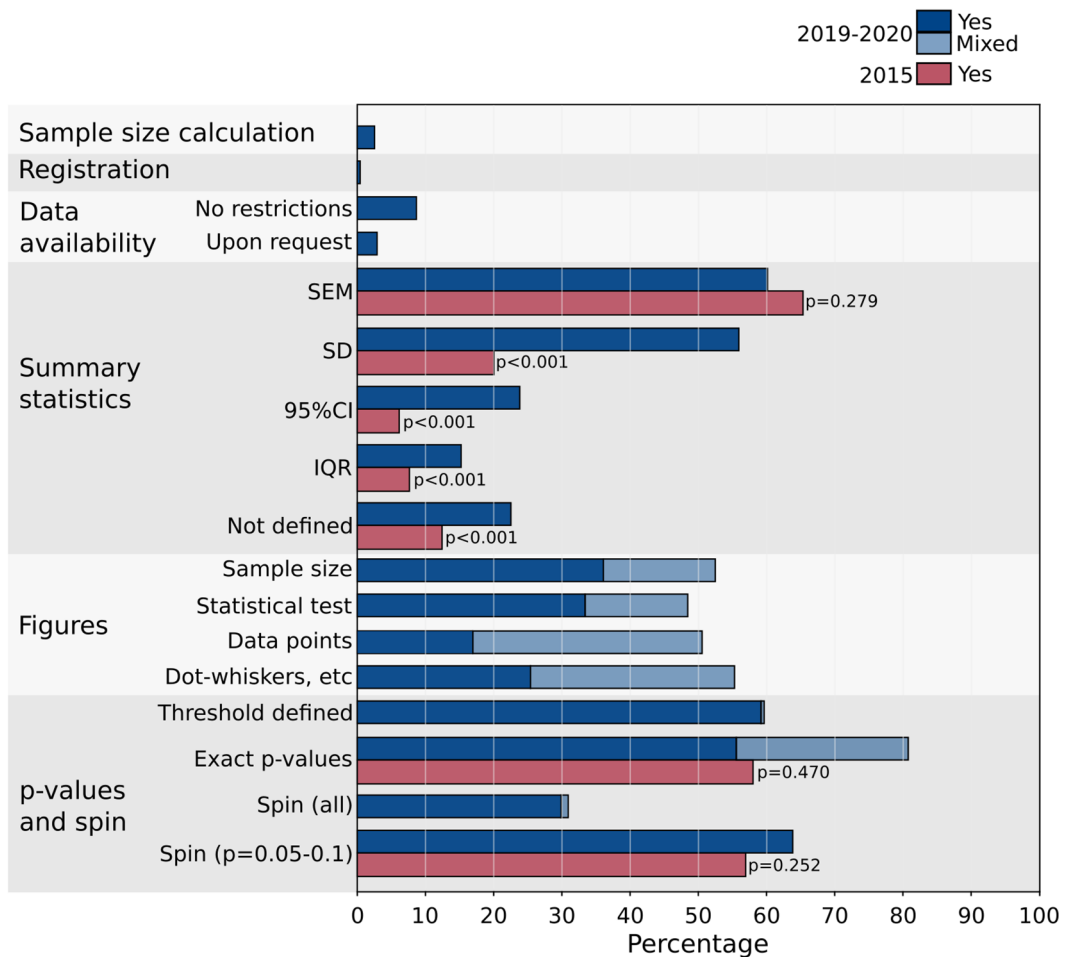
Data were summarised with counts and percentages of papers that fulfilled the scoring criteria for each audit item. As it was not directly relevant to the audit, risk of bias was not assessed. Data processing and analysis were performed in Python (v3.9). Raw data and summary results are provided (underlying data: *Audit Results*). Although this study is not a systematic review, PRISMA guidelines were adhered where possible<sup>42</sup> (extended Data: *PRISMA Checklists*).

**Statistical analyses**

For audit items from the original audit of papers published in 2015 in the Journal of Neurophysiology,<sup>33</sup> we compared the proportions of responses for results from the current audit to those of the previous audit using the `proportions_ztest` function from the Python `statsmodels` package,<sup>43</sup> with the level of significance set at  $\alpha=0.05$ .

**Results**

Summary results for the present audit and the 2015 audit are presented in [Figure 1](#), including the statistical results of the proportions tests. Detailed results from the present and previous audits are presented in [Table 1](#).



**Figure 1. Audit results.** The results of the present audit (2019-2020 papers) are plotted in blue. Values in light blue represent the percentage of papers that adhered to the reporting requirement, but not consistently. Also plotted are the results from the previous audit of 2015 Journal of Neurophysiology papers, including the results from tests of proportions between the 2015 and 2019-2020 results. Detailed results are presented in [Table 1](#). Standard error of the mean (SEM); interquartile range (IQR); standard deviation (SD); confidence interval (CI).

**Table 1. Percentages and count data from current and previous audits.**

	JNP -2020 current study	JNP Héroux 2016	JP and BJP 2007-2015 Diong et al. 2018	NeuRA 2017-2018 Héroux et al. 2022
Sample size calculation	2.7% (15/556)	—	—	13.3% (56/421)
Registered	0.5% (3/580)	—	—	6.2% (28/452)
Data availability <sup>†</sup>	9.0% (52/580)	—	—	6.0% (27/452)
	3.1% (18/580)	—	—	—
Summary statistics	60.2% (343/570)	65.4% (178/278)	80.0% (593/741) <sup>§</sup>	32.4% (138/426)
SEM*	56.0% (319/570)	20.2% (55/278)	—	—
SD	23.9% (131/570)	6.2% (17/278)	—	—
95% CI*	15.3% (87/570)	7.7% (21/278)	—	—
IQR	22.6% (129/570)	12.5% (34/278)	12.9% (97/750) <sup>§</sup>	30.4% (129/424)
Not defined	36.7% (205/559)	—	—	—
Sample size <sup>††</sup>	Mixed:	—	—	—
	Yes:	—	—	—
Statistical test <sup>††</sup>	Yes:	—	—	—
	Mixed:	—	—	—
	Yes:	—	—	—
	Mixed:	—	—	—
Data points	Yes:	—	2.8% (10/356) <sup>¶</sup>	—
	Mixed:	—	—	—
Dot-whisker, etc <sup>†</sup>	Yes:	—	—	—
	Mixed:	—	—	—
p-values and spin	Yes:	—	58.4% (160/274)	62.3% (255/409)
Threshold defined*	Mixed:	—	—	—
	Yes:	58.4% (160/274)	7.4% (52/698) <sup>¶</sup>	57.4% (231/402) <sup>¶</sup>
Exact p-values*	Mixed:	—	—	—
	Yes:	—	—	—
Spin (all)	Yes:	—	—	31.1% (127/409)
	Mixed:	—	—	—
Spin (p=0.05-0.1)	64.0% (171/267)	56.8% (42/74)	60.0% (36/60)	—

Journal of Neurophysiology (JNP); Journal of Physiology (JP); British Journal of Pharmacology (BJP); Neuroscience Research Australia (NeuRA); standard error of the mean (SEM); interquartile range (IQR); standard deviation (SD); confidence interval (CI).

\*Indicates items that were encouraged or, for the SEM, discouraged.

<sup>†</sup>Indicates items that were strongly encouraged.

<sup>††</sup>Indicates items that were required.

<sup>§</sup>Values pooled and averaged across years; applied to summary statistics reported in text and figures.

<sup>¶</sup>Scored 'yes' if item always adhered to; mixed reporting scored as 'no'.

In total, 580 original research papers published in 2019 and 2020 were audited. Few papers included a sample size calculation (2.7%) or indicated that the study protocol was registered prior to the start of the study (0.5%). Similarly, few papers made their data available publicly (9.0%) or upon request (3.1%), despite this practice being strongly encouraged by the updated American Physiological Society Information for Authors.

Overall, 60.2% of papers reported SEM, similar to the 65.4% noted in the original audit of 2015 papers. While the percentage of papers that reported SD, IQR, and 95% CI nearly doubled between 2015 and 2019-2020, so too did the percentage of papers that included undefined measures of variability (12.5% vs 22.6%).

Although it was a required reporting practice, only a third of papers consistently specified the type of statistical tests or the number of samples or animals in their figure legends. Similarly, despite being strongly recommended, only a third of papers consistently plotted their data in a manner that conveyed the variability of the data.

In line with results from two previous audits,<sup>32,34</sup> only 59.3% of papers defined a probability threshold. And finally, in line with results from the previous audit of 2015 papers, only 55.7% of papers consistently reported exact p-values, while 64.0% of papers with p-values between 0.05 and 0.1 reported them as statistical trends.

## Discussion

We provide evidence that publicly highlighting poor reporting practices in a journal does not lead to improvements. Moreover, our results indicate that strongly encouraging or requiring reporting practices does not guarantee these will be adhered to or enforced.

### Interpreting audit results

Compared to the original audit, the majority of poor reporting practices were unchanged or worsened. For example, ~60% of audited papers reported SEM while ~40% of papers failed to specify a threshold value for their statistical tests. Also, p-values between 0.05 to 0.1 were misinterpreted as statistical trends more than 60% of the time, a result that confirms researcher bias and highlights that spin continues to be a major problem in science.<sup>20</sup> Worryingly, the number of papers that included undefined summary statistics increased from 12.5% of papers in 2015 to 22.6% of papers in 2019-2020. Such high rates of spin and poor scientific reporting are worrisome. Overall, results of the present audit show that, at least in this journal, highlighting poor reporting practices does nothing to remedy the problem.

In 2016 and again in 2018, the American Physiological Society introduced several required or strongly encouraged reporting practices. Our results indicate that easy-to-implement required reporting practices were consistently adhered to in only a third of audited papers. Similarly, strongly encouraged reporting practices were consistently adhered to in only 12.0 to 25.0% of audited papers. While the introduction of new reporting requirements to improve the quality, transparency and reproducibility of published papers is well intentioned, our results indicate that they were not universally adopted and enforced at the Journal of Neurophysiology.

### Why does poor research reporting persist?

The above results raise an important question: Who is ultimately responsible for enforcing the reporting requirements outlined in a journal's *Information for Authors*? The obvious answer is the journal itself, its staff and its editors. However, as highlighted by our own work and that of others,<sup>12-23</sup> staff and editors across several journals do not take on this responsibility. While it is possible that editors do not agree with reporting practices required by their journal and thus do not choose to enforce them, a more plausible explanation is that editors, like the rest of us, are overworked.<sup>44</sup> They do not have the time to ensure papers they handle comply with every aspect of their *Information for Authors*. In line with this possibility, personal communication with an Associate Editor at the Journal of Neurophysiology confirms that editors are not required to read in their entirety the papers they handle. A similar comment was shared by an Associate Editor of the European Journal of Applied Physiology, highlighting that this is not an isolated practice. In addition, some editors lack knowledge about reporting guidelines while others fear that authors will not submit to a journal renowned for its strict reporting guidelines.<sup>45</sup> In the same vein, journal staff likely do not have the time (or expertise) to review every paper with a fine tooth comb to ensure they comply with reporting requirements. Some journals, including the Journal of Neurophysiology, use SciScore (<https://sciscore.com/>) to assess and score submitted manuscripts in an attempt to enhance the rigor and reproducibility of published papers. Unfortunately, these tools are far from perfect and their output requires expert review. Moreover, the items assessed by SciScore at the Journal of Neurophysiology do not address any of the items in the present audit. Thus, despite being the obvious choice to enforce their own reporting requirements, many journals are not currently enforcing compliance with their *Information for Authors*.

Who else might be responsible for this enforcement? Reviewers? Reviewers are expected to focus on the science, not compliance with reporting requirements. Having said this, several journals require reviewers to complete a series of

tick-box questions to indicate whether or not the paper they are reviewing complies with key reporting practices. At the Journal of Neurophysiology, reviewers are required to answer the following questions: 'Should any bar graphs be presented as box-whisker plots or dot-whisker plots, instead of bar graphs, to enhance clarity?', 'Do bar graphs show individual data values?', and, with regards to figure legends, 'Do they include details for each individual experiment (sample number (n), sex, statistics test)?'. Based on the results of the present audit, this approach does not seem effective.

What about authors? Authors are expected to read a journal's *Information for Authors* when preparing their manuscript. However, *Information for Authors* can be lengthy, convoluted and difficult to navigate, with each journal having their own bespoke version. Authors may not follow the *Information for Authors* to save time. Since the Journal of Neurophysiology has a rejection rate of ~55%,<sup>46</sup> authors may not feel it is worth their time to comply with all of the requirements and recommendations in the *Information for Authors*. This is especially true when authors can peruse the current issue of the Journal of Neurophysiology and see that reporting requirements are not enforced. Authors already have to deal with high rejection rates, pedantic submission processes and tedious formatting requirements (at submission or acceptance) that amount to free labour for journals and publishers.<sup>47</sup> To pass the responsibility of compliance with *Information for Authors* to the authors would be taking further advantage of the skewed power dynamic that exists between journals, publishers and authors. Having said this, a simple checklist related to compliance with reporting requirements could be required at submission or acceptance. However, someone would still need to verify and enforce results from these checklists. As mentioned above, a similar checklist is already in place at the Journal of Neurophysiology for reviewers, and it appears to be largely ineffective. Thus, it is unclear whether introducing a checklist for authors to complete would fare any better.

### Possible solutions to improve research reporting

A possible solution would be for editors to be assigned a small team of junior editors. These junior editors would assess papers that go out for review for compliance with reporting practices outlined in the *Information for Authors*. Such an approach, if adhered to and properly enforced, could ensure required and recommended reporting practices become the norm rather than the exception. In fact, some American Physiological Society Journals have introduced an Early Career Editorial Fellowship Program.<sup>48</sup> This is a great initiative, and the scope of the program could be broadened to tackle the tasks mentioned above. Unfortunately, few *Information for Authors* address good reporting practices<sup>49</sup>; an issue that would need to be addressed for this solution to work.

Another possible solution would be for journals to employ additional staff with the experience and expertise to review submitted manuscripts. It may be that software tools can be used in the first instance, with journal staff carefully reviewing papers only once they are likely to be accepted. The obvious problem with this solution is that some journals are run on small budgets, while other journals are run by large profit-driven corporations that are unlikely to support the creation of such positions.

In healthcare, published guidelines do not automatically translate into clinical practice.<sup>50</sup> This type of behavioural change must be supported by multiple strategies that target all aspects of a system.<sup>51</sup> A similar approach may be required to change adherence to reporting guidelines in academic journals.

### Limitations

We acknowledge there are limitations to conducting an audit of papers published in a single journal. How generalisable are our results? While the present study can be viewed as a case study, its results are, for the most part, in line with two previous large-scale audits: one that involved over 750 papers published in the Journal of Physiology and the British Journal of Pharmacology,<sup>32</sup> the other that involved nearly 500 papers published by a medical research institute across dozens of journals and disciplines.<sup>34</sup> Our results are also in line with other recent reports of poor scientific reporting,<sup>17–23,52</sup> including an audit where 61% of published COVID-19 studies used incorrect statistical methods.<sup>53</sup> Although the present audit focused on a single journal, we did not set out to target the Journal of Neurophysiology *per se*. Our previous 2015 audit<sup>33</sup> and the introduction of new reporting requirements presented a unique opportunity to answer the research questions tackled here. Whether the present results are applicable to other journals remains to be determined.

To reduce the bias in data extraction to a minimum, it would be preferable if future audits were performed in duplicate. Two investigators would audit each paper and any disagreements would be resolved in discussion with a third investigator.

### Conclusions

Accurate and unbiased scientific reporting is essential. However, to assume authors will always comply with reporting requirements and apply research best practices is idealistic. Similarly, to assume that editors and reviewers – who, at other



times, are also authors – will rise to the occasion without stronger incentives (or compensation) is also unrealistic. Education does not lead to substantial change,<sup>32</sup> nor does updating *Information for Authors* or publicly highlighting poor reporting practices, as this audit illustrates. Novel solutions that acknowledge and address the complexity of the scientific enterprise and the various incentives that are at play are sorely needed.

## Data availability

### Underlying data

Figshare: Poor statistical reporting, inadequate data presentation and spin persist despite Journal awareness and updated *Information for Authors*. <https://doi.org/10.6084/m9.figshare.c.6920920.v1>.<sup>54</sup>

This project contains the following underlying data:

- **Audit results.** Spreadsheet containing all audit data. Also included are summary statistics for each audit item (overall, 2019, 2020).
- **Extract from *Information for Authors*.** Extract from the American Physiological Society *Information for Authors*. Specifically, the sections on 1) Promoting Transparent Reporting in APS Publications to Enhance Data Reproducibility; 2) Data presentation, and 3) Experimental Details to Report in Your Manuscript.

### Extended data

This project contains the following extended data:

- **Scoring manual.** Manual that explains each audit item. Key terms and the scope of each audit item are defined, and possible answers specified. Examples are also provided.
- **PRISMA Checklists.** Completed PRISMA Checklist and PRISMA Abstract Checklist.

Data are available under the terms of the <https://creativecommons.org/licenses/by/4.0/> Creative Commons Attribution 4.0 International license (CC-BY 4.0).

## References

1. Curran-Everett D, Benos DJ: **Guidelines for reporting statistics in journals published by the American Physiological Society.** *Am. J. Physiol. Regul.* 2004; **287**(2): R247–R249. [PubMed Abstract](#) | [Publisher Full Text](#)
2. Lang TA, Altman DG: **Basic statistical reporting for articles published in biomedical journals: The SAMPL Guidelines.** *Science Editors' Handbook.* 2013; pp. 29–32.
3. Wood J, Freemantle N, King M, et al.: **Trap of trends to statistical significance: likelihood of near significant P value becoming more significant with extra data.** *BMJ.* Mar 2014; **348**: g2215. [PubMed Abstract](#) | [Publisher Full Text](#)
4. Button KS, Ioannidis JP, Mokrysz C, et al.: **Power failure: why small sample size undermines the reliability of neuroscience.** *Nat. Rev. Neurosci.* 2013; **14**(5): 365–376. [PubMed Abstract](#) | [Publisher Full Text](#)
5. Halsey LG, Curran-Everett D, Vowler SL, et al.: **The fickle P value generates irreproducible results.** *Nat. Methods.* 2015; **12**(3): 179–185. [PubMed Abstract](#) | [Publisher Full Text](#)
6. Wilkinson MD, Dumontier M, Aalbersberg JJ, et al.: **The FAIR Guiding Principles for scientific data management and stewardship.** *Sci. Data.* 2016; **3**: 160018. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Nguyen TV, Rivadeneira F, Civitelli R: **New guidelines for data reporting and statistical analysis: Helping authors with transparency and rigor in research.** *JBMR.* 2019; **34**(11): 1981–1984. [PubMed Abstract](#) | [Publisher Full Text](#)
8. Dirnagl U: **The p value wars (again).** *Eur. J. Nucl. Med. Mol. Imaging.* 2019; **46**(12): 2421–2423. [PubMed Abstract](#) | [Publisher Full Text](#)
9. Drummond G: **A world beyond P: policies, strategies, tactics and advice.** *Exp. Physiol.* Jan 2020; **105**(1): 13–16. [PubMed Abstract](#) | [Publisher Full Text](#)
10. Harris JK, Combs TB, Johnson KJ, et al.: **Three changes public health scientists can make to help build a culture of reproducible research.** *Public Health Rep.* 2019; **134**(2): 109–111. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Prager EM, Chambers KE, Plotkin JL, et al.: **Improving transparency and scientific rigor in academic publishing.** *Brain Behav.* Jan 2019; **9**(1): e01141. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Nagele P: **Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals.** *BJA.* 2003; **90**(4): 514–516. [PubMed Abstract](#) | [Publisher Full Text](#)
13. Curran-Everett D, Benos DJ: **Guidelines for reporting statistics in journals published by the American Physiological Society: the sequel.** *Adv. Physiol. Educ.* Dec 2007; **31**(4): 295–298. [PubMed Abstract](#) | [Publisher Full Text](#)
14. Boutron I, Dutton S, Ravaud P, et al.: **Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes.** *JAMA.* 2010; **303**(20): 2058–2064. [Publisher Full Text](#)
15. Chiu K, Grundy Q, Bero L: **'Spin' in published biomedical literature: A methodological systematic review.** *PLoS Biol.* 2017; **15**(9): e2002173–16. 15457885. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
16. Weissgerber TL, Garcia-Valencia O, Garovic VD, et al.: **Why we need to report more than 'Data were Analyzed by t-tests or ANOVA'.** *elife.* Dec 2018; **7**. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

17. Hildebrandt T, Prenoveau JM: **Rigor and reproducibility for data analysis and design in the behavioral sciences.** *Behav. Res. Ther.* Mar 2020; **126**: 103552.  
[PubMed Abstract](#) | [Publisher Full Text](#)
18. Yanni W, Zhou C, Wang R, et al.: **Statistical reporting in nursing research: Addressing a common error in reporting of p values (p =.000).** *J. Nurs. Scholarsh.* 2020; **52**(6): 688–695.  
[PubMed Abstract](#) | [Publisher Full Text](#)
19. Nieminen P, Uribe SE: **The quality of statistical reporting and data presentation in predatory dental journals was lower than in non-predatory journals.** *Entropy.* Apr 2021; **23**(4).  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Bero L, Lawrence R, Leslie L, et al.: **Cross-sectional study of preprints and final journal publications from COVID-19 studies: discrepancies in results reporting and spin in interpretation.** *BMJ Open.* 2021; **11**(7): e051821.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Schaafsma H, Laasanen H, Twynstra J, et al.: **A review of statistical reporting in dietetics research (2010-2019): How is a Canadian journal doing?** *Can. J. Diet. Pract. Res.* Jun 2021; **82**(2): 59–67.  
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Sauerbrei W, Haeussler T, Balmford J, et al.: **Structured reporting to improve transparency of analyses in prognostic marker studies.** *BMC Med.* May 2022; **20**(1): 184.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. White NM, Balasubramaniam T, Nayak R, et al.: **0.05 was considered statistically significant and other cut-and-paste statistical methods.** *PLoS One.* 2022; **17**(3): e0264360.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Drummond GB, Paterson DJ, McLoughlin P, et al.: **Statistics: all together now, one step at a time.** *J. Physiol.* 2011; **589**: 1859.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Drummond GB, Vowler SL: **Type I: families, planning and errors.** *J. Physiol.* 2012; **590**(20): 4971–4974.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Drummond GB, Tom BD: **Statistics, probability, significance, likelihood: words mean what we define them to mean.** *J. Physiol.* 2011; **589**: 3901–3904.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Drummond GB, Vowler SL: **Not different is not the same as the same: how can we tell?** *J. Physiol.* 2012; **590**(21): 5257–5260.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Drummond GB, Tom BD: **How can we tell if frogs jump further?** *J. Physiol.* 2011; **589**: 3409–3413.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Drummond GB, Vowler SL: **Data interpretation: using probability.** *J. Physiol.* 2011; **589**(Pt 10): 2433–2435.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Drummond GB, Tom BD: **Presenting data: can you follow a recipe?** *J. Physiol.* 2011; **589**: 5007–5011.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Drummond GB, Vowler SL: **Show the data, don't conceal them.** *J. Physiol.* 2011; **589**(Pt 8): 1861–1863.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. Diong J, Butler AA, Gandevia SC, et al.: **Poor statistical reporting, inadequate data presentation and spin persist despite editorial advice.** *PLoS One.* 2018; **13**: e0202121.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Héroux ME: **Inadequate reporting of statistical results.** *J. Neurophysiol.* 2016; **116**(3): 1536–1537.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Héroux ME, Butler AA, Cashin AG, et al.: **Quality Output Checklist and Content Assessment (QuOCCA): a new tool for assessing research quality and reproducibility.** *BMJ Open.* 2022; **12**(9): e060976.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Ioannidis JP: **Why most published research findings are false.** *PLoS Med.* 2005; **2**(8): e124.  
[PubMed Abstract](#) | [Publisher Full Text](#)
36. Calin-Jageman RJ, Cumming G: **Estimation for better inference in neuroscience.** *eNeuro.* 2019; **6**(4): ENEURO.0205–19.2019.
37. Kerr NL: **HARKing: hypothesizing after the results are known.** *Personal. Soc. Psychol. Rev.* 1998; **2**(3): 196–217.  
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Dwan K, Altman DG, Arnaiz JA, et al.: **Systematic review of the empirical evidence of study publication bias and outcome reporting bias.** *PLoS One.* 2008; **3**(8): e3081.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
39. Munafò MR, Nosek BA, Bishop DVM, et al.: **A manifesto for reproducible science.** *Nat. Hum. Behav.* 2017; **1**: 0021.  
[Publisher Full Text](#)
40. Hardwicke TE, Ioannidis JPA: **Mapping the universe of registered reports.** *Nat. Hum. Behav.* 2018; **2**(11): 793–796.  
[PubMed Abstract](#) | [Publisher Full Text](#)
41. Nosek BA, Beck ED, Campbell L, et al.: **Preregistration is hard, and worthwhile.** *Trends Cogn. Sci.* 2019; **23**(10): 815–818.  
[PubMed Abstract](#) | [Publisher Full Text](#)
42. Page MJ, McKenzie JE, Bossuyt PM, et al.: **The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.** *BMJ.* Mar 2021; **372**: n71.  
[Publisher Full Text](#)
43. Seabold S, Perktold J: **statsmodels: Econometric and statistical modeling with python.** *9th Python in Science Conference.* 2010.
44. Vandenberg JP: **STREGA, STROBE, STARD, SQUIRE, MOOSE, PRISMA, GNOSIS, TREND, ORION, COREQ, QUOROM, REMARK... and CONSORT: for whom does the guideline toll?** *J. Clin. Epidemiol.* 2009; **62**(6): 594–596.  
[PubMed Abstract](#) | [Publisher Full Text](#)
45. Grindlay DJC, Dean RS, Christopher MM, et al.: **A survey of the awareness, knowledge, policies and views of veterinary journal editors-in-chief on reporting guidelines for publication of research.** *BMC Vet. Res.* 2014; **10**(1): 10.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
46. Journal of Neurophysiology: **Journal info - journal of neurophysiology.** accessed 10/5/2023.  
[Reference Source](#)
47. LeBlanc AG, Barnes JD, Saunders TJ, et al.: **Scientific sinkhole: The pernicious price of formatting.** *PLoS One.* 2019; **14**(9): e0223116.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. American Journal of Physiology: Renal. **Ajp-renal early career editorial fellowship program.** accessed 10/5/2023.  
[Reference Source](#)
49. Diong J, Bye E, Djajadikarta Z, et al.: **Encouraging responsible reporting practices in the Instructions to Authors of neuroscience and physiology journals: There is room to improve.** *PLoS One.* 2023; **18**(3): e0283753.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
50. Sarkies MN, Jones LK, Gidding SS, et al.: **Improving clinical practice guidelines with implementation science.** *Nat. Rev. Cardiol.* 2022; **19**(1): 3–4.  
[Publisher Full Text](#)
51. Grol R, Grimshaw J: **From best evidence to best practice: Effective implementation of change in patients' care.** *Lancet.* 2003; **362**(9391): 1225–1230.  
[Publisher Full Text](#)
52. Amiri M, Deckert M, Michel MC, et al.: **Statistical inference in abstracts of three influential clinical pharmacology journals analyzed using a text-mining algorithm.** *Br. J. Clin. Pharmacol.* Mar 2021; **87**: 4173–4182.  
[Publisher Full Text](#)
53. Ordak M: **COVID-19 research: quality of biostatistics.** *Arch. Med. Sci.* 2022; **18**(1): 257–259.  
[PubMed Abstract](#) | [Publisher Full Text](#)
54. Héroux ME, Diong J, Bye E, et al.: **Poor statistical reporting, inadequate data presentation and spin persist despite journal awareness and updated information for authors.** *figshare. Collection.*  
[Publisher Full Text](#)

# Open Peer Review

Current Peer Review Status: ? ✓ ? ✓

---

## Version 1

Reviewer Report 28 February 2024

<https://doi.org/10.5256/f1000research.156438.r236603>

© 2024 McCrum C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



### Christopher McCrum

Maastricht University Medical Centre, Maastricht, The Netherlands

The manuscript presents the results of an audit of specific reporting practices in recent issues of a journal and compares these with previous audits, in order to evaluate if changes in author guidelines and requirements have influenced practice. This is an interesting and important manuscript and while only conducted for one journal, the results are in line with many previous audits at other journals. The discussion highlights some of the important issues in attempting to improve these practices. The manuscript is clearly written and reported and the associated data and resources are accessible (except for the PRISMA material mentioned). I have no major comments that the authors need to address. I have a few general comments for the authors to consider.

Some phrasing in the manuscript indicates that causality might be determinable with the current auditing approach (e.g., "We provide evidence that publicly highlighting poor reporting practices in a journal does not lead to improvements.") but I think only association could be determined by observing changes in reporting following journal guideline changes. Consider rephrasing some sections of the discussion on this point.

The approach and many results are comparable to results on sample size calculation reporting in two recent studies that I have been involved in. In those papers, journals that explicitly required justification of the sample size used also only saw adherence in the ~30-40% range (1, 2). The current manuscript clearly confirms the findings that including requirements in author guidelines are not sufficient to drastically improve reporting. Therefore, I wonder if one additional/expanded recommendation in the discussion that could be made is the inclusion of a specific question in the peer review process (either for editors or reviewers), expanding on the current checkbox approach that the authors discuss, on whether or not the manuscript adheres to a stated journal requirement and where in the article this information is provided. Acceptance of the manuscript for publication could be dependent on a positive response on this item (I think often not the case). This would not require any additional staff or budget and would not necessarily require additional expertise from the reviewers, since they only have to be able to identify the presence of the information, not its validity. I think the discussed concern about journal appearing to "strict" in

their requirements lacks evidence, since some of the most prestigious medical journals tend to have extremely specific and detailed requirements for their manuscripts (not just statistical reporting) and still attract submissions. I understand it is not the aim of the current manuscript to provide solutions, but the authors could expand on their thoughts for future solutions to explore.

### References

1. Twomey R, Yingling V, Warne J, Schneider C, et al.: Nature of Our Literature. *Communications in Kinesiology*. 2021; **1** (3). [Publisher Full Text](#)
2. McCrum C, van Beek J, Schumacher C, Janssen S, et al.: Sample size justifications in Gait & Posture. *Gait Posture*. 2022; **92**: 333-337 [PubMed Abstract](#) | [Publisher Full Text](#)

### Is the background of the case's history and progression described in sufficient detail?

Yes

### Is the work clearly and accurately presented and does it cite the current literature?

Partly

### If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

### Are all the source data underlying the results available to ensure full reproducibility?

Yes

### Are the conclusions drawn adequately supported by the results?

Partly

### Is the case presented with sufficient detail to be useful for teaching or other practitioners?

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Human Movement Science, Meta-Science

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 09 February 2024

<https://doi.org/10.5256/f1000research.156438.r229985>

© 2024 F. D. Carneiro C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Clarissa F. D. Carneiro**

Berlin Institute of Health at Charité, QUEST Center for Responsible Research,, Berlin, Germany

The paper presents the results of an assessment of the quality of statistical reporting in publications from the Journal of Neurophysiology (2019-2020). The paper is very concise and direct in its presentation of the study, what makes it easy and pleasant to read. The down side of this conciseness is that we depend on consulting the supplementary materials to understand some details of the table (e.g., why do the denominators change in different items) and some bits of the discussion are quite vague. I list a few suggestions and open questions below, but how much to deepen the discussion and description of the methods and data in the main manuscript text can be a matter of personal preference. In any case, the data and other additional materials available are very well organized and easy to consult and re-use.

- The main concern I have is with the use of causal language. As this is an observational study, causality cannot be inferred but is implied in some sections of the text. To highlight just one example, the first sentence of the discussion (“We provide evidence that publicly highlighting poor reporting practices in a journal does not lead to improvements.”) implies a causality link but the correlation would be better described as “We provide evidence that publicly highlighting poor reporting practices in a journal is not associated with improvements.”.
- The main open question I have is if the creation and updating of the *Information to authors* by the publisher was accompanied by similar updates in the guides/instructions to reviewers or any training material to editors. Reviewers and editors can have lots of different interpretations about their roles (e.g.: Glonti *et al.*, 2019<sup>1</sup>; Glonti and Hren, 2018<sup>2</sup>), and reviewers assessing papers for different journals or publishers may not be aware of specific editorial policies or requirements.
- The different ways that spin is referred to throughout the text can be confusing for readers who are not familiar with the concept. Spin can be much more complex and difficult to detect/assess than simply describing a p-value close to 0.05 as a trend of significance, and this is only superficially implied throughout this paper. If the authors prefer not to go into detail of definitions within the current paper, listing references 14 and 15 next to the definitions of spin could help guide the reader.
- A trivial question that emerged as reading the paper was if the Journal of Neurophysiology is published by the American Physiological Society. This could be an obvious link to physiologists and can easily be solved with a quick search online, but might be worth explicitly stating in the introduction for a better logic flow.
- Regarding the use of checklists as mentioned in the discussion, there is further evidence suggesting they are not very effective in improving quality of reporting which would make the argument against checklists for authors stronger (e.g. Hair *et al.*, 2019<sup>3</sup>). In addition, these tasks if unchecked can quickly compound with other redundant or meaningless ones contributing to a negative impact and/or perception of responsible research practices (e.g., Cochran *et al.*, 2024<sup>4</sup>).
- I personally agree with the suggestion that the journal staff should be primarily responsible for reviewing quality of reporting, but also understand the limitations described. Do the authors believe the problems of small budgets and profit-driven motivations could be reduced in large scientific society publishers, like the APS? It could be an interesting addition to the discussion.

- Regarding behavioral changes, I'd also be curious to read more on the authors views on approaches from clinical practice that could be translated to the publication system. There are other models of behavioral change being proposed for open/responsible research practices that might also be worth discussing (e.g.: Norris and O'Connor, 2019<sup>5</sup>; Robson et al., 2021<sup>6</sup>; <https://www.ncbi.nlm.nih.gov/books/NBK556627/>).
- One final small question, when assessing data availability: if a repository link or ID is given, did the assessors check that it worked? Or just presenting a link was sufficient to score "yes"? This is a potential limitation (e.g.: Federer (2022<sup>7</sup>)).

## References

1. Glonti K, Cauchi D, Cobo E, Boutron I, et al.: A scoping review on the roles and tasks of peer reviewers in the manuscript review process in biomedical journals. *BMC Med.* 2019; **17** (1): 118 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Glonti K, Hren D: Editors' perspectives on the peer-review process in biomedical journals: protocol for a qualitative study. *BMJ Open.* 2018; **8** (10): e020568 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Hair K, Macleod MR, Sena ES, IICARus Collaboration: A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus). *Res Integr Peer Rev.* 2019; **4**: 12 [PubMed Abstract](#) | [Publisher Full Text](#)
4. Cochran JD, Leathers TA, Maldosevic E, Siejda KW, et al.: Cell cycle specific, differentially tagged ribosomal proteins to measure phase specific transcriptomes from asynchronously cycling cells. *Sci Rep.* 2024; **14** (1): 1623 [PubMed Abstract](#) | [Publisher Full Text](#)
5. Norris E, O'Connor D: Science as behaviour: Using a behaviour change approach to increase uptake of open science. *Psychology & Health.* 2019; **34** (12): 1397-1406 [Publisher Full Text](#)
6. Robson S, Baum M, Beaudry J, Beitner J, et al.: Promoting Open Science: A Holistic Approach to Changing Behaviour. *Collabra: Psychology.* 2021; **7** (1). [Publisher Full Text](#)
7. Federer LM: Long-term availability of data associated with articles in PLOS ONE. *PLoS One.* 2022; **17** (8): e0272845 [PubMed Abstract](#) | [Publisher Full Text](#)

**Is the background of the case's history and progression described in sufficient detail?**

Yes

**Is the work clearly and accurately presented and does it cite the current literature?**

Partly

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Partly

**Is the case presented with sufficient detail to be useful for teaching or other practitioners?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** meta-research, scholarly publication, open, responsible and reproducible research practices

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Reviewer Report 09 February 2024

<https://doi.org/10.5256/f1000research.156438.r229987>

© 2024 Dunleavy D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Daniel J. Dunleavy** 

Florida State University, Tallahassee, Florida, USA

The authors have written a case study about the statistical reporting practices of articles published in the *Journal of Neurophysiology (JNP)* for the years 2019 and 2020. The study supports a wide ranging body of work, which finds suboptimal, if not erroneous, reporting of statistical methods and results across the biomedical and social sciences. (Pocock *et al.*, 1987<sup>1</sup>; Nuijten *et al.*, 2016<sup>2</sup>; Dunleavy, 2020<sup>3</sup>; Ziliak, 2004<sup>4</sup>).

While there have been calls and initiatives to improve the reporting of quantitative studies within the field of physiology - for example, editorials published by the *Journal of Physiology* and *British Journal of Pharmacology*; efforts by the American Physiological Society to revise its "Information for Authors" pages - little is known about their impact.

To explore this, the authors audited 580 papers published in JNP, published after changes to the journal's Information for Authors page had been put into effect. Many of the reporting practices examined remained unchanged or worsened between the two timepoints.

Overall, I found the paper to be of interest, its methods relatively sound, and its conclusions reasonable.

I hope the following comments, questions, and suggestions help strengthen the overall quality of the manuscript. Additionally, please be advised that any and all references included in this review are suggestive and not meant to be viewed as compulsory.

#### **Data Availability and Integrity:**

1a. I can confirm that I was able to access the spreadsheet containing audit results for each article.

Further, I selected a handful of results and was able to identify and locate the corresponding article using the "paper" variable.

1b. After opening several papers, I was able to verify a handful of the coded results.

2a. I was also able to locate and understand the scoring manual and locate a copy of the "Information for Authors" page.

2b. However, I was not able to locate the PRISMA Checklist associated with the current study.

### Reporting of Methods and Results:

3. The authors indicated that there were 580 published papers audited in the study. Personally, I'd like to see more context about the total number of papers published in the journal, for the two years (e.g., including a breakdown of articles excluded and their type - editorials, commentaries, errata. etc.).

4. The authors state, "Two investigators would audit each paper and any disagreements would be resolved in discussion with a third investigator." I'd like to see more information about how common disagreements were, even if they were able to be fully resolved. This would permit the reader a clearer understanding about whether certain variables/items were more difficult to code than others.

### Miscellaneous:

5. The authors do a good job acknowledging the limitations of their study - primarily that it focuses on a single journal, during a short period of time. However, the title of this case study does not make this clear. If possible, I'd recommend changing the title of the paper to something more descriptive, like:

"Poor statistical reporting, inadequate data presentation and spin persist: A study of papers published in the Journal of Neurophysiology, 2019-2020"

6. To the extent that preprints are used in this field, the authors might consider exploring what changes are made between preprint version of manuscripts and their published versions. This might give some indication into how peer review and editorial oversight at JNP impacts the content and reporting of papers. Some interesting work is being done by Mario Malički and colleagues (Malički *et al.*, 2024<sup>5</sup>)

### References

1. Pocock SJ, Hughes MD, Lee RJ: Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med*. 1987; **317** (7): 426-32 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Nuijten MB, Hartgerink CH, van Assen MA, Epskamp S, et al.: The prevalence of statistical reporting errors in psychology (1985-2013). *Behav Res Methods*. 2016; **48** (4): 1205-1226 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Dunleavy D: Appraising Contemporary Social Work Research: Meta-Research on Statistical Reporting, Statistical Power, and Evidential Value. 2020. [Publisher Full Text](#)
4. Ziliak S, McCloskey D: Size matters: the standard error of regressions in the American Economic



Review. *The Journal of Socio-Economics*. 2004; **33** (5): 527-546 [Publisher Full Text](#)

5. Malički M, Jerončić A, ter Riet G, Bouter L, et al.: Comparison of Reports of Epidemiology Studies Posted as bioRxiv Preprints and Published in Peer Reviewed Journals. *International Congress on Peer Review and Scientific Publication*. 2024. [Reference Source](#)

**Is the background of the case's history and progression described in sufficient detail?**

Yes

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Partly

**Is the case presented with sufficient detail to be useful for teaching or other practitioners?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** I am an academic social worker by training (PhD) with a record of publishing on issues related to meta-science, open science, peer review, and other issues in scholarly publishing.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 14 December 2023

<https://doi.org/10.5256/f1000research.156438.r226880>

© 2023 Kahlert U. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Ulf Kahlert**

University Clinic for General-, Visceral-, Vascular- and Transplantation Surgery, Medical Faculty University Hospital, Magdeburg,, Germany

Héroux and colleagues conducted a tedious work. Comparing pre- and post audit papers published by an investigated journal incorporates comparing of changes in behavior of authors

and the journal. This setup is ideal for comparison and worth investigating. The presented results from this sector of research continues to be worrisome.

The required reporting items as defined in the paper differ from the journal information for authors. Journal states that the statistical test and number of animals/models per group shall be stated in the material and methods section. Although I agree to Héroux and colleagues that stating these details in the figure legend is necessary to “sufficiently describe the figure on their own, without reference to the main text (as stated in the Information for authors of the journal homepage), the scoring for reporting completeness might be different. On the journal homepage, for the figure legends, regarding statistics it says: “Define all statistical symbols and abbreviations”.

I think the title of the paper shall indicate to the main conclusion drawn: introduction of recommendations/guidelines have not sufficient penetration to all authors. This may draw more attention, as lot of people focus on recommendations - which is good - but the impact might be overestimated.

Suggested aspects to be changed: What responsibility could the authors' home institution play? Shall publication with using a given affiliations imply control measures of written results before submission. Also grant agencies: shall there be a paragraph to provide published evidence that a lab of a given applicant enforces measures / adhere to complete reporting? The society debates about sustainability. Only reproducible science has high chance for sustaining.

**Is the background of the case's history and progression described in sufficient detail?**

Yes

**Is the work clearly and accurately presented and does it cite the current literature?**

Partly

**If applicable, is the statistical analysis and its interpretation appropriate?**

I cannot comment. A qualified statistician is required.

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Is the case presented with sufficient detail to be useful for teaching or other practitioners?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** cancer research

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have**

**significant reservations, as outlined above.**

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**