Published in final edited form as:

Ann Intern Med. 2018 June 05; 168(11): 832–833. doi:10.7326/M17-2863.

Statistical code for clinical research papers in a high-impact specialist medical journal

Melissa Assel, M.S.¹, Andrew J. Vickers, PhD¹

¹Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Background

It is widely accepted that statistical analyses should be implemented by writing good quality code in a professional statistical package such as R, SAS or Stata. Good code ensures reproducibility, reduces error, and provides auditable documentation of the analyses underpinning research results. There have been several recent efforts to encourage archiving of code corresponding to published papers^{1–5}, on the grounds that doing so improves transparency. Such efforts have focused on areas such as neuroscience or bioinformatics, which are highly dependent on computationally intensive analyses.

Objective

To examine how often authors used statistical code for clinical research papers published in a high-impact specialty journal, and to determine the quality of this code.

Methods and Findings

In mid-2016, we added to the online submission system for *European Urology* a question regarding whether authors used statistical code and, if so, whether they would be willing to submit it were their paper to be accepted. In August 2017, we reviewed 314 papers subsequently accepted to the journal. Authors of 40 papers reported that they used statistical code. Authors archived the code with the journal for 18 of these papers, with the remaining 32 declining to do so.

We randomly selected and reviewed 50 papers where authors had reported no code. Of these 50, 35 presented no statistics (e.g. a narrative review of the literature) or only trivial analyses (e.g. a single survival curve). The remaining 15 included substantive analyses, such as large

Corresponding Author Andrew J. Vickers, Memorial Sloan Kettering Cancer Center, 485 Lexington Avenue, New York, NY 10017 USA, vickersa@mskcc.org, (P): 646-888-8233.

Author Contributions

Concept and design: Vickers

Acquisition, analysis or interpretation of data: Assel, Vickers, Catto, Pierce, Redley, Le

Drafting of the manuscript: Assel, Vickers

Critical revision of the manuscript for important intellectual content: Assel, Vickers

Statistical analysis: Assel, Vickers

Administrative, technical, or material support: Catto, Pierce, Redley, Le

There are no conflicts of interest

numbers of regression models, graphs or time-to-event statistics. We contacted the corresponding authors for these 15 papers; 8 told us that they did not use code but 7 responded that they had indeed used code and that their initial response was erroneous. In 6 of these 7 cases, the authors refused to submit their code to the journal.

We then examined the all code sets received, excluding code associated with 3 papers submitted by authors trained in our group. Most of the code had little or no annotation and extensive repetition. For half of the papers, the reviewed code included no formatting for presentation (Table 1).

Discussion

No statistical code was used for more than a third of papers published in a high-impact specialist medical journal that included non-trial statistical analyses. Not a single set of code managed to score even moderately on three basic and widely accepted software criteria. This is not a superficial problem. For instance, failure to include code that formats numerical output increases the risk of transcription errors; repeated code can lead to inconsistent analyses.

We have three recommendations. First, software practices and principles should become a core part of biostatistics curricula, irrespective of both degree (undergraduate or post-graduate) and subject (biostatistics, public health or epidemiology). Given that students will have to write code when they come to perform analyses as practicing investigators, we find it difficult to understand why few degree programs in quantitative medical science teach good coding practice. Second, there should be intramural peer review of statistical code. Colleagues should routinely share code with one another for the purposes of constructive criticism just as they share drafts of scientific papers. Third, code associated with published research should be archived. Doing so would not only improve transparency and reproducibility but would also help to ensure that investigators would write better quality code. One investigator we contacted told us that he was unwilling to archive his code with the journal because he had not “made any effort to make it ... usable by others”, with the result that much of it was “dirty”. Our view is that the value of well-written code goes far beyond the cosmetic, and that “dirty” code may well lead to scientific errors.

Failure to use statistical programming code, or writing of poor quality code, importantly threatens the validity of scientific findings. We urge the medical research community to take immediate remedial action.

Acknowledgement

Funding

This work was supported by a National Institutes of Health/National Cancer Institute Cancer Center Support Grant to MSKCC (grant number P30-CA008748).

References

1. Eglen SJ, Marwick B, Halchenko YO, et al. Toward standard practices for sharing computer code and programs in neuroscience. *Nat Neurosci* 2017;20(6):770–773. [PubMed: 28542156]

2. O'Neill K, Brinkman RR. Publishing code is essential for reproducible flow cytometry bioinformatics. *Cytometry Part A* 2016;89(1):10–11.
3. Herbert R, Elkins M. Publishing code: an initiative to enhance transparency of data analyses reported in *Journal of Physiotherapy*. *Journal of Physiotherapy* 2017;63(3):129–130. [PubMed: 28652079]
4. Bernard C Editorial: Code Case - Investigating Transparency and Reproducibility. *eneuro* 2017;4(4).
5. <https://www.nature.com/news/code-share-1.16232>. Accessed September 2017.

Table 1.

Assessment of code of 16 published articles.

Domain	Score	N
Are codes well annotated?	Extensive annotation such that one could substantially recreate the code using the annotations.	0
	Moderate annotation such that one could recreate of at least some code on the basis of the annotations.	2
	Little or no annotation.	14
Does the code avoid repetition?	Good use of loops and macros such that there is little or only trivial repeated code.	0
	Moderate amount of repeated code (about 10 lines or fewer).	0
	No use of loops or macros and extensive repetition of code (10 lines or more).	15
	Not applicable; no opportunity for repeated code.	1
Does the code include formatting for presentation?	Markup language or other techniques used to generate formatted output provided for all or almost all of a paper's results.	3
	Markup language or other techniques used to generate formatted output need non-trivial amendment or many results not formatted.	5
	No formatted output.	8