Reproducibility of COVID-19 research papers on bioRxiv and medRxiv*

TBD

Annie Collins, University of Toronto

12 February 2021

Abstract

We create a dataset of all the papers published on bioRxiv and medRxiv between X and Y. We extract the text from these papers and parse them for keywords to do with the availability of data and scripts underpinning the paper. We find that X per cent of papers have X. Our paper demonstrates the need for Y.

1 Introduction

...

We follow Weissgerber et al. (2021)...

The remainder of this paper is structured as follows...

2 Data

We construct our dataset by first using the APIs of bioRxiv and medRxiv to download any paper that:

3 Model

In paper we run our analysis in `R` (R Core Team 2020).

4 Results

5 Discussion

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

*Code and data are available at: https://github.com/anniecollins/reproducibility.

Appendix

References

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Weissgerber, Tracey, Nico Riedel, Halil Kilicoglu, Cyril Labbé, Peter Eckmann, Gerben Ter Riet, Jennifer Byrne, et al. 2021. "Automated Screening of COVID-19 Preprints: Can We Help Authors to Improve Transparency and Reproducibility?" *Nature Medicine* 27 (1): 6–7.