

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

TBD

Annie Collins<sup>†</sup>

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

Scientists use open repositories of papers to more quickly disseminate their research than is possible in traditional journals. These repositories, such as arxiv, bioRxiv, and medRxiv, are a critical component of science and many results build on the work published there. So it is important that the results that are published are credible. These repositories are not peer-reviewed, and, in general, anyone with appropriate academic credentials can submit a paper.

While neither peer-review nor credentials are a panacea nor a guarantee of quality, given the importance of these repositories, it is important that scientists impose on themselves various standards for their results. Following Weissgerber et al. (2021) we examine papers about COVID-19 published to bioRxiv and medRxiv during 2020. We search for markers of open science and reproducibility, such as X, Y, and Z.

We find that A, B, and C.

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

We run our analysis in R (R Core Team 2020).

---

\*We thank CANSSI... Code and data are available at: <https://github.com/anniecollins/reproducibility>.

<sup>†</sup>University of Toronto

## **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.

## 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.