

medrxivr: Accessing and searching medRxiv and bioRxiv preprint data in R

Luke A McGuinness^{1, 2} and Lena Schmidt¹

¹ Department of Population Health Science, University of Bristol ² MRC Intergrative Epidemiology Unit, University of Bristol

DOI: [10.21105/joss.02651](https://doi.org/10.21105/joss.02651)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Daniel S. Katz](#) ↗

Reviewers:

- [@danielskatz](#)

Submitted: 08 September 2020

Published: 09 October 2020

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

An increasingly important source of health-related bibliographic content are preprints: preliminary versions of research articles that have yet to undergo peer review. The two preprint repositories most relevant to health-related sciences are [medRxiv](#) and [bioRxiv](#), both of which are operated by the Cold Spring Harbor Laboratory, a not-for-profit research and educational institution (Rawlinson & Bloom, 2019).

The goal of the `medrxivr` R package is two-fold. In the first instance, it provides programmatic access to the Cold Spring Harbour Laboratory (CSHL) API, allowing users to download medRxiv and bioRxiv preprint metadata (e.g., title, abstract, author list.) This functionality will be of interest to anyone who wishes to import medRxiv and/or bioRxiv preprint metadata into R, for example to explore the distribution of preprints by subject area or by publication year. Examples of this type of usage have already been reported (e.g., by Brierley, 2020).

In the second instance, the package provides functions that allow users to search the downloaded preprint metadata for relevant preprints using complex search strings, including functionality such as search term truncation, Boolean operators (AND, OR, NOT), and term proximity. Helper functions are provided that allow users to export the results of their search to a .bib file for import into a reference manager (e.g., Zotero) and to download the full-text PDFs of preprints matching their search. This aspect of the package will be more relevant to systematic reviewers, health librarians and others performing literature searches, allowing them to perform and document transparent and reproducible searches in these important evidence sources.

Acknowledgements

We acknowledge funding from NIHR (LAM through NIHR Doctoral Research Fellowship (DRF-2018-11-ST2-048), and LS through NIHR Systematic Reviews Fellowship (RM-SR-2017-09-028)). LAM is a member of the MRC Integrative Epidemiology Unit at the University of Bristol. The views expressed in this article are those of the authors and do not necessarily represent those of the NHS, the NIHR, MRC, or the Department of Health and Social Care.

References

- Brierley, L. (2020). The role of research preprints in the academic response to the COVID-19 epidemic. *Authorea*. doi:[10.22541/au.158516578.89167184](https://doi.org/10.22541/au.158516578.89167184)
- Rawlinson, C., & Bloom, T. (2019). New preprint server for medical research. *BMJ*, 365. doi:[10.1136/bmj.l2301](https://doi.org/10.1136/bmj.l2301)