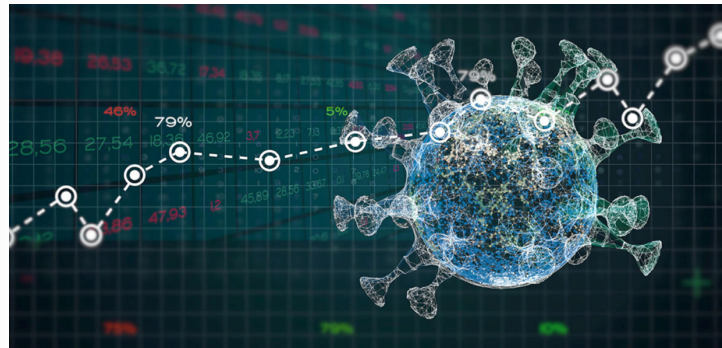
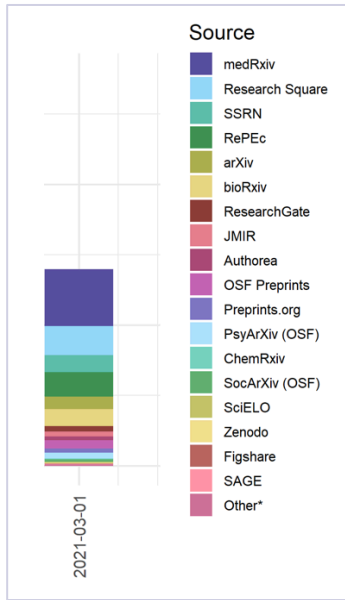


April 14, 2021

Preprint–Publication Linking applied to COVID19



1. Problem Statement



https://github.com/nicholasmfraser/covid19_preprints

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

CSH Cold Spring Harbor Laboratory BMJ Yale

Causal Impact of Masks, Policies, Behavior on Early Covid-19 Pandemic in the U.S

Victor Chernozhukov, Hiroyuki Kasahara, Paul Schrimpf
doi: <https://doi.org/10.1101/2020.05.27.20115139>
Now published in *Journal of Econometrics* doi: [10.1016/j.jeconom.2020.09.003](https://doi.org/10.1016/j.jeconom.2020.09.003)

Abstract Full Text Info/History Metrics

ARTICLE INFORMATION

doi: <https://doi.org/10.1101/2020.05.27.20115139>
History September 12, 2020.

ARTICLE VERSIONS

Version 1 (May 29, 2020 - 03:17).
Version 2 (May 30, 2020 - 19:48).
Version 3 (June 30, 2020 - 15:30).
Version 4 (July 9, 2020 - 13:48).
Version 5 (July 13, 2020 - 11:53).
You are viewing **Version 6**, the most recent version of this article.

RESEARCH ARTICLE

The effect of online multimedia psychoeducational interventions on the perceived stress and resilience of hospitalized patients with COVID-19: a quasi-experimental study

Maryam Shaygan, Zahra Yazdani, Adib Vailbeigi

DOI: [10.21203/rs.3.rs-36980/v1](https://doi.org/10.21203/rs.3.rs-36980/v1) Download PDF

LICENSE: © ⓘ This work is licensed under a CC BY 4.0 License. [Read Full License](#)

View the published article at [BMC Psychiatry](#).

Research Square
BADGES
Prescreen

PEER REVIEW TIMELINE
CURRENT STATUS: PUBLISHED
Version 2
Posted 28 Oct, 2020
View this version
Version 1
Posted 17 Jul, 2020
No comments provided

- Readers Researchers, systematic reviewers, journalists, the public
- Biomed preprint servers 56+ ([Kirkham et al., 2020](#) and <https://asapbio.org/preprint-servers>)
- Change in evidence Preprint $v_1 \neq v_N$ and preprint \neq publication (30%, [Oikonomidi et al., 2020](#))
- Preprint–publication linking Neither immediate nor comprehensive (37% only, [Cabanac et al., 2021](#))

Question: How can my preprint on medRxiv be linked to the journal version when it is published?

Answer: medRxiv will usually automatically add a link to the published version within approximately 2 weeks of journal publication. The corresponding author should receive an email from medRxiv requesting confirmation of the link. On rare occasions a match is not made because the title or authors have changed. Please wait 2–3 weeks before contacting medRxiv staff should the link not appear.

<https://www.medrxiv.org/about/FAQ>

Day-to-day Preprint–Publication Link Discovery

- **Input**
 - The most up-to-date scientific literature (17k new publications a day)
 - One preprint
- **Output** A list of candidate publications sorted by decreasing relevance
- **Constraints**
 - Effective: Matches publications ASAP (early view...)
 - Efficient: Fast 'enough' (depends on the volume to handle)
- **Applications**
 - Add publication link on preprint servers
 - Refresh living systematic reviews to revise expert knowledge
 - In bibliometric studies: merge citations to a preprint and its linked publications

Preprint–Publication Linker

- **Source** Crossref bibliographic data
 - Public domain (CC0)
 - Most publication DOIs are Crossref-minted
 - Latest changes reflected immediately through the API (no data stored locally)
- **Method** Match and filter Crossref records on selected criteria:
 - Timeline compatibility
 - Title fuzzy overlap
 - Byline head: common first author (ORCID or identity)
 - Doctype
- **Pros** Outsourced indexing+searching, up-to-date data source, small footprint, free of charge
Flexible criteria
- **Cons** Retrieves publications with DOIs and minted by Crossref (~~no DOI~~ or Datacite)
3 minutes per preprint with the public API = small volumes
- **Users** Cochrane epidemiologists for the [Living systematic review of COVID19 trials](#)
Daily assessments and updates of <https://www.irit.fr/~Guillaume.Cabanac/covid19-preprint-tracker>

Benchmarking Linkers

■ Sensitivity/specificity analysis on a 343-preprint test collection

		True status of a preprint		Total
		Published preprint	Unpublished preprint	
Preprint servers	Published preprint	48 True positive	0 False positive (Type I Error)	48
	Unpublished preprint	73 False negative (Type II Error)	222 True negative	295
Total		121	222	343

Measure	%
Sensitivity	39.7
Specificity	100.0
Positive predicted value (PPV)	100.0
Negative predicted value (NPV)	75.3
Accuracy	78.7

Fig. 5 Evaluation of the **preprint servers** on the 343 preprint reference set, as of 23 October 2020

		True status of a preprint		Total
		Published preprint	Unpublished preprint	
Linker's output	Published preprint	110 True positive	18 False positive (Type I Error)	128
	Unpublished preprint	11 False negative (Type II Error)	204 True negative	215
Total		121	222	343

Measure	%
Sensitivity	90.9
Specificity	91.9
Positive predicted value (PPV)	85.9
Negative predicted value (NPV)	94.9
Accuracy	91.5

Fig. 6 Evaluation of the **preprint–publication linker** on the 343 preprint reference set, as of 23 October 2020

⚠ Testing recall is hard work: regular time-consuming manual searches to find all existing links
Focusing on precision is easier...

■ Precision-oriented benchmark

- Standard evaluation procedure of the *Text REtrieval Conference* series (NIST)
- Precision@1: Average of the precisions at rank 1 for each preprint (1 if the top-ranked publication is relevant else 0)
- Reciprocal rank: Average of the reciprocal ranks for each preprint (1/rank of the topmost relevant result or 0)
- Test collections:
 - a. 245 preprint-publication links validated by a Cochrane epidemiologist
 - b. Thousands of pairs registered at Crossref and medRxiv