

Preprint servers – the purgatory of peer-reviewed publications?

Juliette C Thompson, Aditi Hombali, Eric J Manalastas, David A Scott || Visible Analytics Ltd.

Juliette.Thompson@VisibleAnalytics.co.uk || Aditi.Hombali@VisibleAnalytics.co.uk || Eric.Manalastas@VisibleAnalytics.co.uk || David.Scott@VisibleAnalytics.co.uk

INTRODUCTION

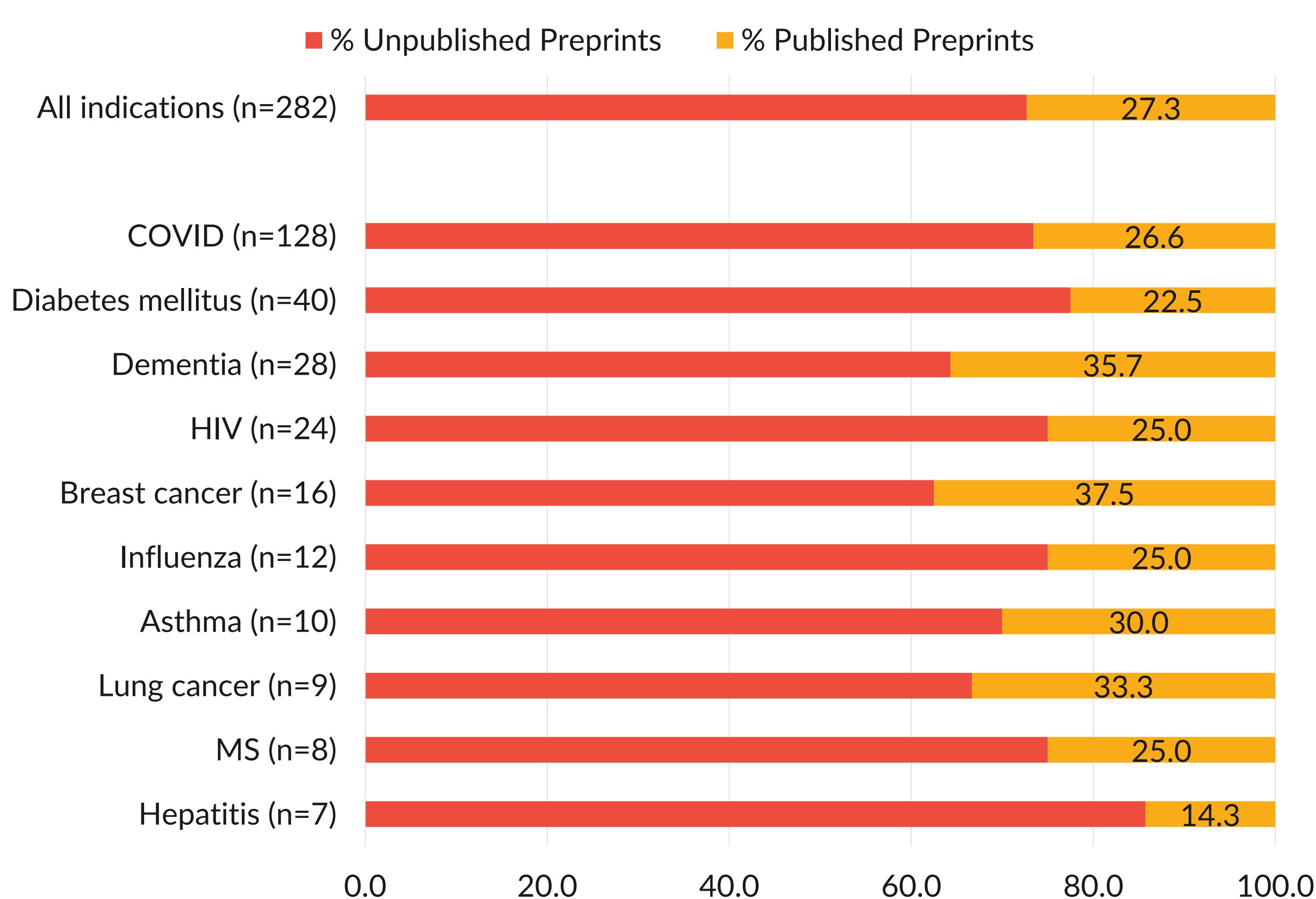
- Research is increasingly being published via preprint servers to make results available in advance of peer-reviewed publication. These publications are not subject to referee or peer-review.
- Databases such as PubMed and Embase also include publications from preprint servers such as medRxiv and bioRxiv; as such systematic literature review (SLR) searches are increasingly likely to also capture preprint publications
- There are no recommendations on whether it is appropriate to include preprints in SLRs and quantitative evidence synthesis, therefore we sought to understand the conversion rates for clinical trial preprints to peer-reviewed publications to determine the scale of the potential issue.

METHODS

- Ten disease indications were randomly selected to cover a broad array of literature including:
 - Infectious diseases (COVID, human immunodeficiency virus [HIV], Influenza, Hepatitis)
 - Oncology (Breast cancer, Lung cancer)
 - Respiratory (Asthma)
 - Endocrinology (Diabetes mellitus)
 - Autoimmune (Multiple Sclerosis [MS])
- A search was conducted in Embase.com to identify randomised clinical trial preprints published in January 2023
- A subsequent search was conducted to determine whether the preprint had been published by 31-May-2024 by searching for title matches between and peer-reviewed publications

RESULTS

- Across the ten case study diseases, 282 preprints were published in January 2023, ranging from 7 to 128 preprints depending on indication
- By 31-May-2024, 22/282 (27.3%) of the preprints had an associated peer-reviewed publication
- Hepatitis had both the lowest number of preprints and lowest conversion to peer-reviewed publications (1/7, 14.3%). The greatest conversion rate was observed in breast cancer where 6/16 (37.5%) preprints were peer-review published
- In indications with the largest number of preprints (COVID-19 [34/128, 26.6%], diabetes mellitus [9/40, 22.5%], dementia [10/28, 35.7%], HIV [6/24, 25.0%]) conversion rates ranged from 22.5% to 35.7%



REFERENCES

- Baldwin M. Peer review. Encyclopedia of the History of Science. (2019).
- Baxt WG, Waechterle JF, Berlin JA, Callahan ML. Who reviews the reviewers? Feasibility of using a fictitious manuscript to evaluate peer reviewer performance. *Ann Emerg Med* 1998;32(3 Pt 1):310-7.
- Bonnechere B. Preprints in Medicine: Useful or Harmful? *Front Med (Lausanne)* 2020;7:579100.
- Chen R, Desai NR, Ross JS, et al. Publication and reporting of clinical trial results: cross sectional analysis across academic medical centers. *BMJ* 2016;352:i637.
- Godlee F, Gale CR, Martyn CN. Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial. *Jama* 1998;280(3):237-40.
- Mastroianni A. The rise and fall of peer review. *Experimental History*; (2022).
- Sampson C. Why I'm quitting peer review. *The Academic Health Economists' Blog*; (2023).
- Schroter S, Black N, Evans S, Godlee F, Osorio L, Smith R. What errors do peer reviewers detect, and does training improve their ability to detect them? *J R Soc Med* 2008;101(10):507-14.

For example, the search line to identify preprints in HIV:

'human immunodeficiency virus' AND [preprint]/lim AND [01-01-2023]/sd NOT [01-02-2023]/sd

A search of individual titles of preprints would either return 1 result, where the preprint was still the only record available, or 2 results where a peer-reviewed publication had subsequently become available

DISCUSSION

Limitations

- We relied on the peer-reviewed publication retaining exactly the same title as was used in the preprint disclosure. It is likely that some authors make changes to the titles of articles between disclosures, and thus possible we over-estimate the number of preprints which have not been published
- The time-frame which we allowed (~16 months) for a peer-reviewed article to be published could be too short a period for a publication to realistically appear. Research indicates that only 28.6% of clinical trials have published results within 24 months of their primary completion date (as recorded on clinicaltrials.gov) (Chen 2016)

Is the peer-review process on the brink?

- The peer-review process as we know it started to materialise in the 1950s before coalescing into what we see now by the mid-1970s (Baldwin 2019). Despite general acceptance there have, and remain, questions over its effectiveness and longevity (Mastroianni 2022; Sampson 2023).
- The purpose of peer-review is to prevent bad research from being published, however several studies indicate that only 25-30% of major flaws are captured (Baxt 1998; Godlee 1998; Schroter 2008). Furthermore, corrective mechanisms in place (errata and retractions) are inadequately and inconsistently implemented meaning that once published research generally remain so{one of our retraction papers}
- The use of preprint servers can offer a valuable rapid dissemination of research with initial aim of garnering comments from peers on articles prior to submission to a more traditional peer-review journal (Bonnechere 2020)
- However, the apparent low rate of subsequent peer-reviewed publications over a year after preprint dissemination does beg three questions:
 - Are authors submitting articles they deposit to preprint servers to peer-reviewed journals' and they are simply not being accepted?
 - Are authors seeing a preprint server as a final repository for their research?
 - If the answers to Q1 and Q2 are yes – then what do you do with this research?
- In the framework of SLR methodology, at present there is no explicit guidance on how to consider preprints. Consideration of grey literature, typically conference disclosures, is recommended, therefore perhaps preprints should be considered in this context. If there is an increasing movement towards preprint servers acting as the terminus of research, it will certainly require contemplation