The Early Access Effect in bioRxiv Preprints

Nicholas Fraser[1] (0000-0002-7582-6339), Fakhri Momeni[2] (0000-0002-5572-575X), Philipp Mayr[2] (0000-0002-6656-1658), Isabella Peters[1,3] (0000-0001-5840-0806)

1. ZBW – Leibniz Information Centre for Economics, Kiel, Germany
2. GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany
3. University of Kiel, Kiel, Germany

Contact:
Nicholas Fraser
n.fraser@zbw.eu

1 Introduction

- **Preprints** are versions of scientific articles that have not yet been accepted for publication in a peer-reviewed journal, and are an important component of modern open science practices.
- A **citation advantage** is predicted for articles deposited as preprints, driven by **increased accessibility** and **earlier availability**, or by factors related to authorship such as **self-selection** of high-quality articles to deposit.
- We aim to investigate the **citation and altmetric impact advantage** for articles deposited to **bioRxiv**, a preprint server for the biological sciences. bioRxiv was launched in November 2013 and now includes over 40,000 preprints.
- Our **research questions** are:
  1. How do citation counts differ between articles deposited to bioRxiv and those not deposited?
  2. How do altmetric indicators differ between articles deposited to bioRxiv and those not deposited?
  3. What additional factors may confound the measured citation and/or altmetric differential?

2 Methods

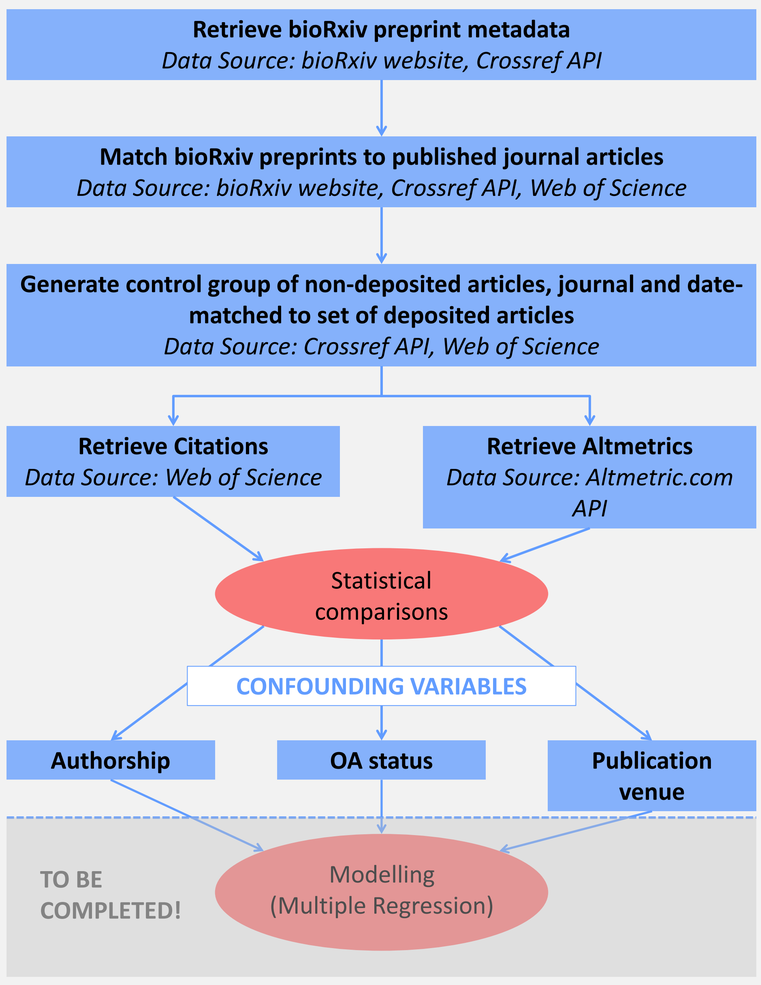


Figure 1: Schematic overview of methodology and major data sources used in this study.

3 Results

Figure 2: bioRxiv usage and publication outcomes, 2013-2017

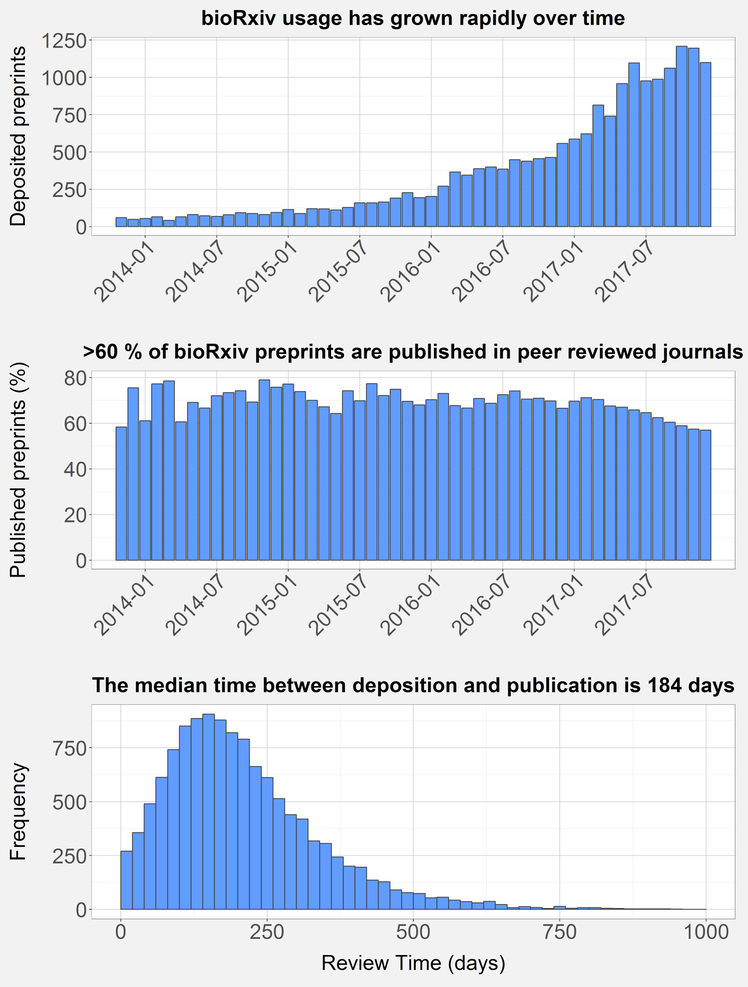


Figure 3: Citation rates of bioRxiv-deposited and non-deposited articles

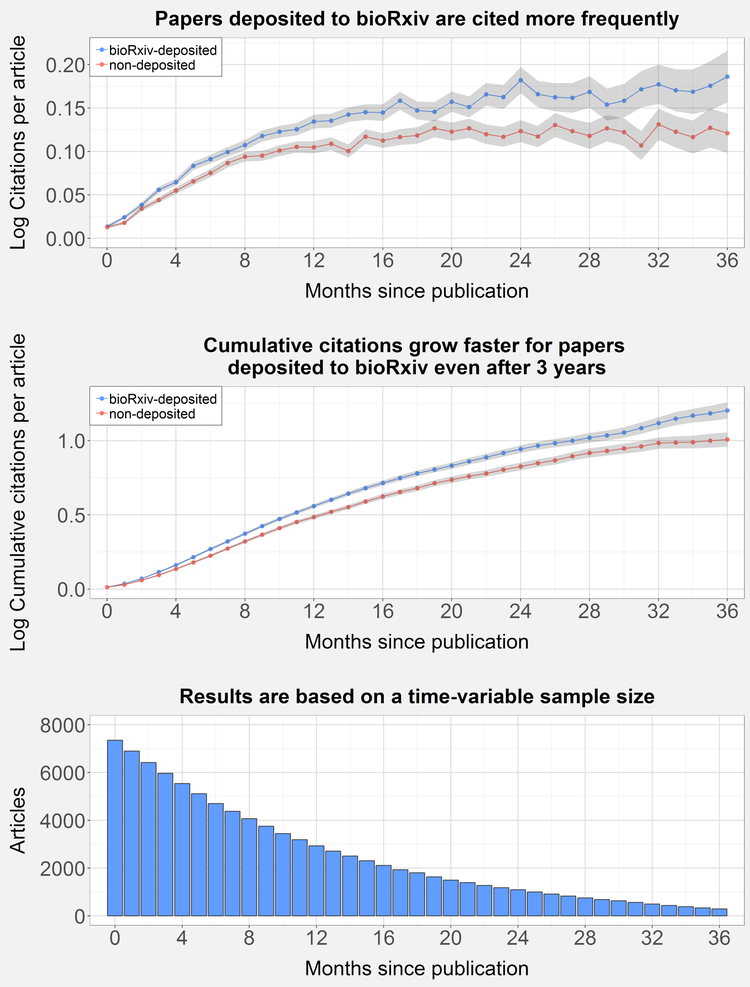


Figure 4: Altmetric counts of bioRxiv-deposited and non-deposited articles

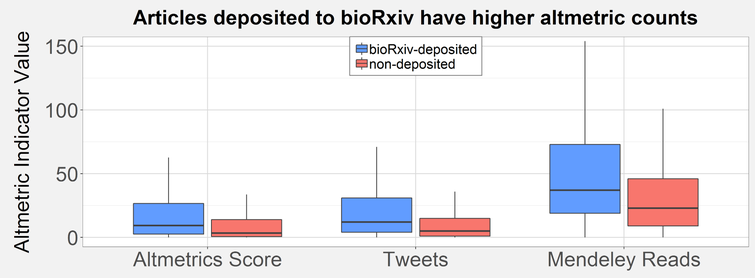
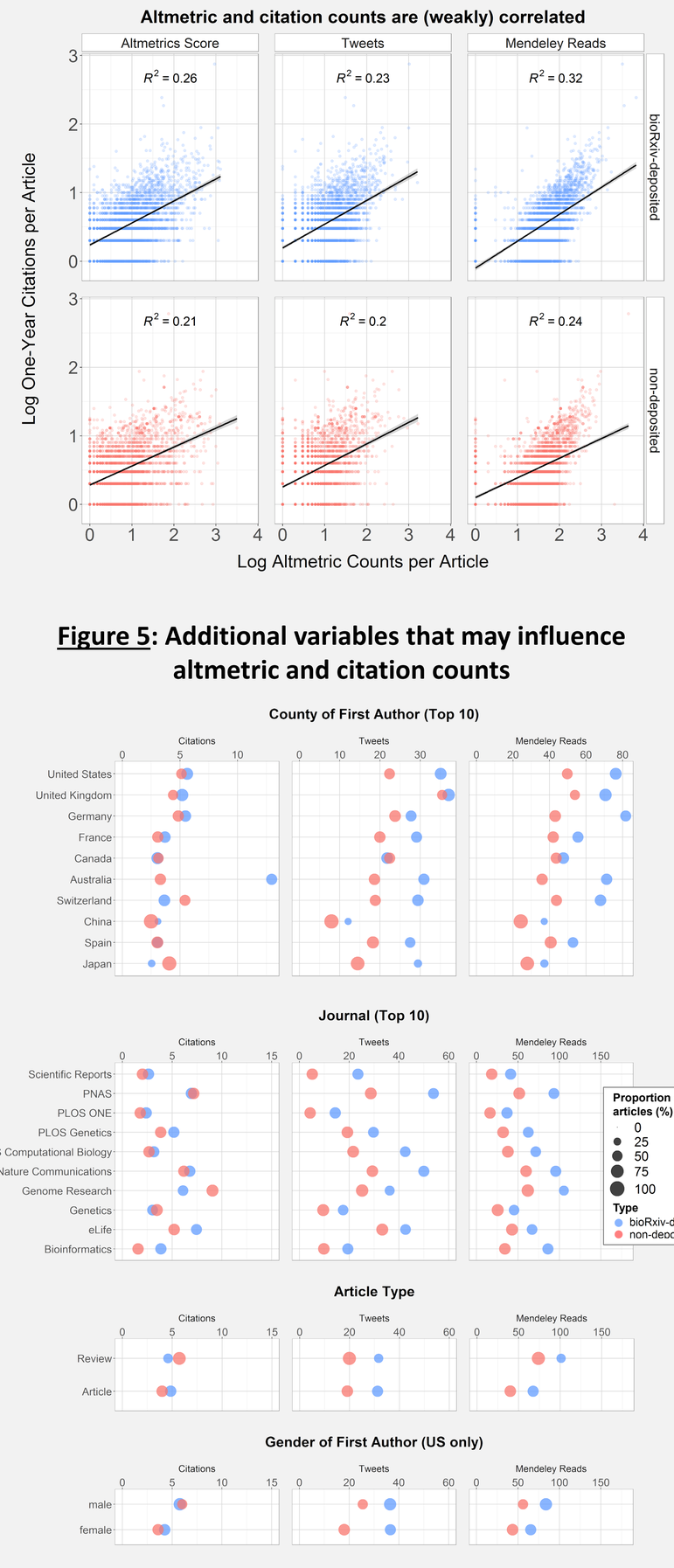


Figure 5: Additional variables that may influence altmetric and citation counts



4 Discussion and Future Work

These initial results show that citations counts and altmetric indicators are significantly greater for articles deposited as bioRxiv preprints versus those that are not. However, we have identified a number of structural and author-specific variables that may confound these observations. Future work will therefore focus upon:

- Identifying potential confounders and data sources for their analysis (to name a few: publication venue, article age, number of co-authors, author countries, author seniority, author gender, article type and journal category).
- Multiple regression analysis to determine the influence of these variables

5 Acknowledgements

This work was funded by BMBF project OASE, grant number 01PU17005A. More information on OASE can be found at http://www.zbw.eu/de/forschung/web-science/oase/ and https://osf.io/ncxh4.