

Citations to preprints in graduate theses at Concordia University: Preliminary results

Krista Alexander
Science Librarian

Katharine Hall
Science Librarian

Melissa Rivosecchi
Head, Vanier Library



LIBRARY

April 24, 2026

Overview

- Rationale and research questions
- Methodology and data collection
- Challenges
- Preliminary results
- Next steps

Project rationale:

Why preprints?

- The COVID-19 pandemic saw a rapid uptake in the use of preprints¹.
- Prior to 2020, preprint use was an established practice in computer science and physics, with a slower uptake in other disciplines.
- Currently we are seeing an increased visibility of preprints, as some large databases, like Scopus and Web of Science, are incorporating preprints into their search corpus and results.

[1] Watson C. Rise of the preprint: how rapid data sharing during COVID-19 has changed science forever. Nature Medicine. 2022;28(1):2-5. doi:[10.1038/s41591-021-01654-6](https://doi.org/10.1038/s41591-021-01654-6)

Research Questions

1. Are graduate students at Concordia University citing preprints in their theses and dissertations?
2. Has the use of preprint literature in graduate student theses changed over time?
3. Is there a difference between levels of preprint citation in Masters compared to PhD theses?
4. Do students acknowledge a source as a preprint either in the text of their thesis or in the reference list?

Methodology:

Thesis sample

- Masters and PhD theses deposited in Spectrum (Concordia's IR)
- Ten-year period: 2015-2024
- Departments of:
 - Biology
 - Chemistry & Biochemistry
 - Health, Kinesiology & Applied Physiology (formerly Exercise Science)
 - Physics
 - Psychology
- Theses with an embargo date post December 2025 were excluded

Methodology:

Preprint identification

- Identified a list of 49 preprint servers and academic websites (eg ResearchGate) that host preprints relevant to our selected disciplines.
- Using that list, we identified 64 search terms
- Searched across all theses and dissertations using Adobe Acrobat's advanced search function
- Reviewed search hits to identify preprints

Search terms	Server
14293	ScienceOpen
21203	Research Square
31124	Advance (Sage preprint)
32388	Qeios
Advance	Advance (Sage preprint)
AIJR	AIJR Preprints
Arabxiv	Arabxiv
ARPHA	ARPHA Preprints
Authorea	Authorea
BECPrep	MitoFit Preprints Archive
Beil	Beilstein Archive
bxiv	
Cell Press	Cell Press Sneak peek
cern	CERN document server
ChinaXiv	ChinaXiv
cogprints	Cogprints
Easy Chair	EasyChair preprints
easychair	
ESS Open	ESSOAr
essoar	
f1000 Res	F1000 Research
f1000res	

Methodology:

Data Collection

For each search hit, we recorded:

- Thesis Year
- Department
- Degree granted
- Thesis structure (Manuscript vs Traditional)
- Number of citations in ref list (if using a style with numbered references)
- Preprint Server
- Preprint Year
- Number of in-text citations to preprint
- Is item identified as preprint/not peer reviewed in text/ref list when cited?
- Preprint citation (as cited in thesis)

Methodological Challenges

- What counts as a preprint?
 - Versioning
 - Open review platforms
 - Manuscript theses/dissertations chapters as preprints
- Permalink/doi in citation points to different source than citation text
- False preprints

Methodological Challenges: Versioning [preprint vs. post-print]

Sample reference list citation:

Wu, H.; Noé, F. Variational Approach for Learning Markov Processes from Time Series Data. 2017.

<https://doi.org/10.48550/ARXIV.1707.04659>.

Submission history

From: Hao Wu [[view email](#)]

[v1] Fri, 14 Jul 2017 23:07:32 UTC (686 KB)

[v2] Mon, 11 Dec 2017 21:20:38 UTC (2,361 KB)

[v3] Fri, 16 Aug 2019 01:52:07 UTC (4,035 KB)

[Submitted on 14 Jul 2017 (this version), latest version 16 Aug 2019 (v3)]

Variational approach for learning Markov processes from time series data

[Submitted on 14 Jul 2017 (v1), revised 11 Dec 2017 (this version, v2), latest version 16 Aug 2019 (v3)]

Variational approach for learning Markov processes from time series data

Methodological Challenges: Open peer review platforms

Sample reference list citation:

J.A. Gustavsen, S. Pai, R.
Isserlin, et al. Rcy3: Network
biology using cytoscape from
within R. F1000Research,
8:1774, 2019.

Open Peer Review			
Reviewer Status		✓✓✓	ⓘ
Reviewer Reports			
	<i>Invited Reviewers</i>		
	1	2	3
Version 3 (revision) 04 Dec 19			
Version 2 (revision) 27 Nov 19	✓ read		✓ read
	↑		
Version 1 18 Oct 19	✓ read	✓ read	

Methodological Challenges: False preprints

False preprint citations:

- Citation mentions "preprint from..." or "preprint at..." but link to published article
- Citation mentions arXiv or bioRxiv but link to published article

113. Frémont,

Biology

114. Simon, C

cytokine

115. Stoten, C

cell divis

<https://doi.org/10.1016/j.semcd.2017.08.035>

116. Schiel, J

mediate c

117. Basant, A

Biology

 View PDF

[Download full issue](#)



Seminars in Cell & Developmental Biology

Volume 74, February 2018, Pages 50-65



Review

ESCRT-dependent control of membrane remodelling during cell division

[Caroline Louise Stoten](#), [Jeremy Graham Carlton](#)  

[Show more](#) 

 Add to Mendeley  Share  Cite

<https://doi-org.lib-ezproxy.concordia.ca/10.1016/j.semcd.2017.08.035> 

[Get rights and content](#) 

Under a Creative Commons [license](#) 

 Open access

Limitations of methodology

Potentially missed hits if hyperlinks were embedded in the text

- [261] C. Degrande, *Automatic evaluation of UV and R2 terms for beyond the Standard Model Lagrangians: a proof-of-principle*, *Comput. Phys. Commun.* **197** (2015) 239–262, [[1406.3030](#)].
- [262] T. Hahn, *Generating Feynman diagrams and amplitudes with FeynArts 3*, *Comput. Phys. Commun.* **140** (2001) 418–431, [[hep-ph/0012260](#)].
- [263] J. Anderson et al., *Snowmass Energy Frontier Simulations*, in *Proceedings, Community Summer Study 2013: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*, 2013. [1309.1057](#).
- [264] A. Avetisyan et al., *Methods and Results for Standard Model Event Generation at $\sqrt{s} = 14$ TeV, 33 TeV and 100 TeV Proton Colliders (A Snowmass Whitepaper)*, in *Proceedings, Community Summer Study 2013: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*, 2013. [1308.1636](#).



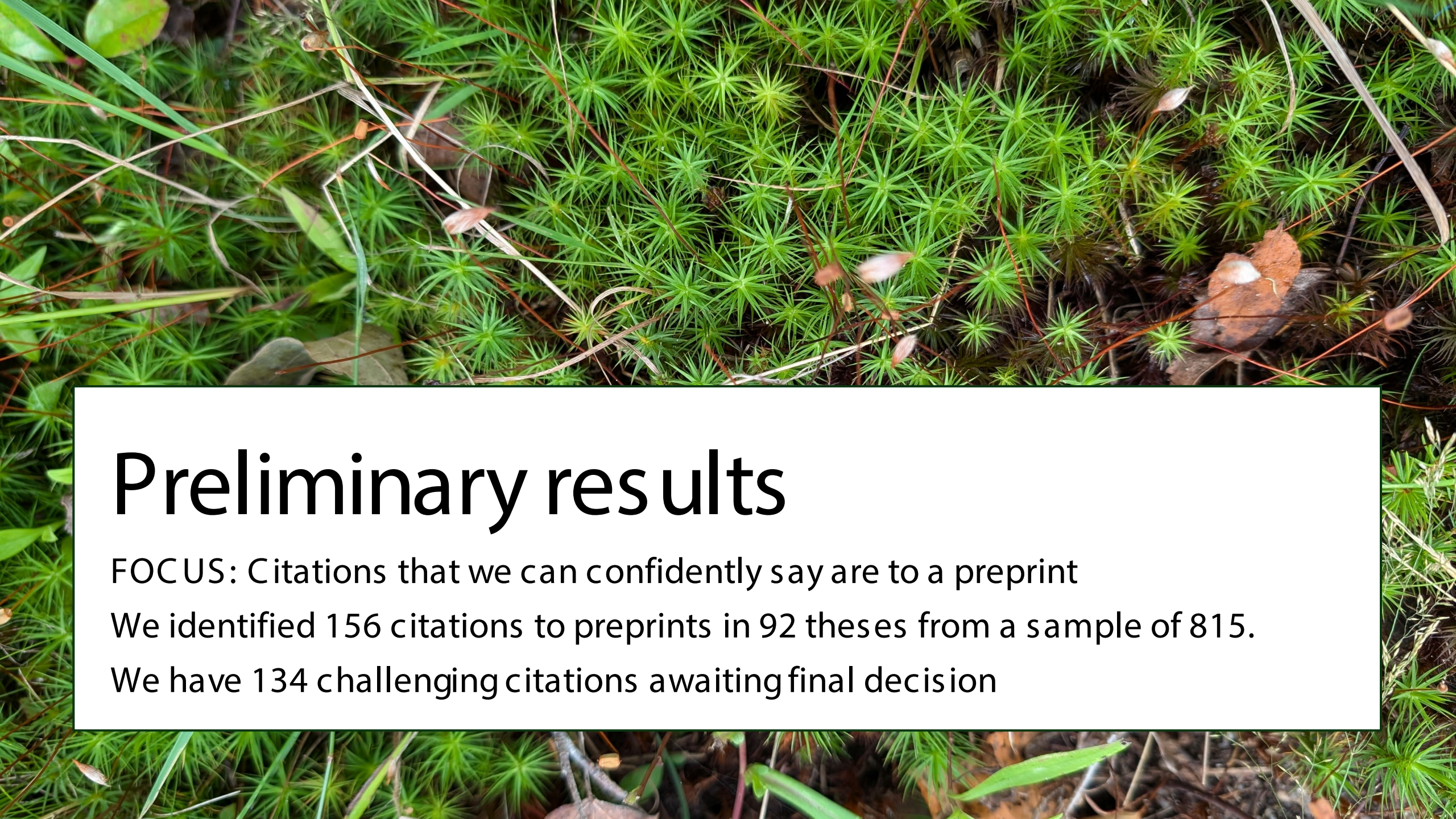
Limitations of methodology : Numbered ref lists

Sample reference list citation:

67. Setty M, Tadmor MD, Reich-Zeliger S, et al. Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat Biotechnol*. 2016;34(6):637-645. doi:10.1038/nbt.3569
68. Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell developmental trajectories. *bioRxiv*. Published online January 1, 2017:110668. doi:10.1101/110668
69. Rizvi AH, Camara PG, Kandror EK, et al. Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. *Nat Biotechnol*. 2017;35(6):551-560. doi:10.1038/nbt.3854

Corresponding in-text citation:

This additional data at both the DNA and mRNA levels will allow us to move towards fuller pseudotime trajectories⁶⁵⁻⁷², a common powerful approach in single cell studies, that

A close-up photograph of a forest floor covered in vibrant green moss. Interspersed among the moss are several dry, brown leaves and thin, light-colored twigs, creating a textured and naturalistic background.

Preliminary results

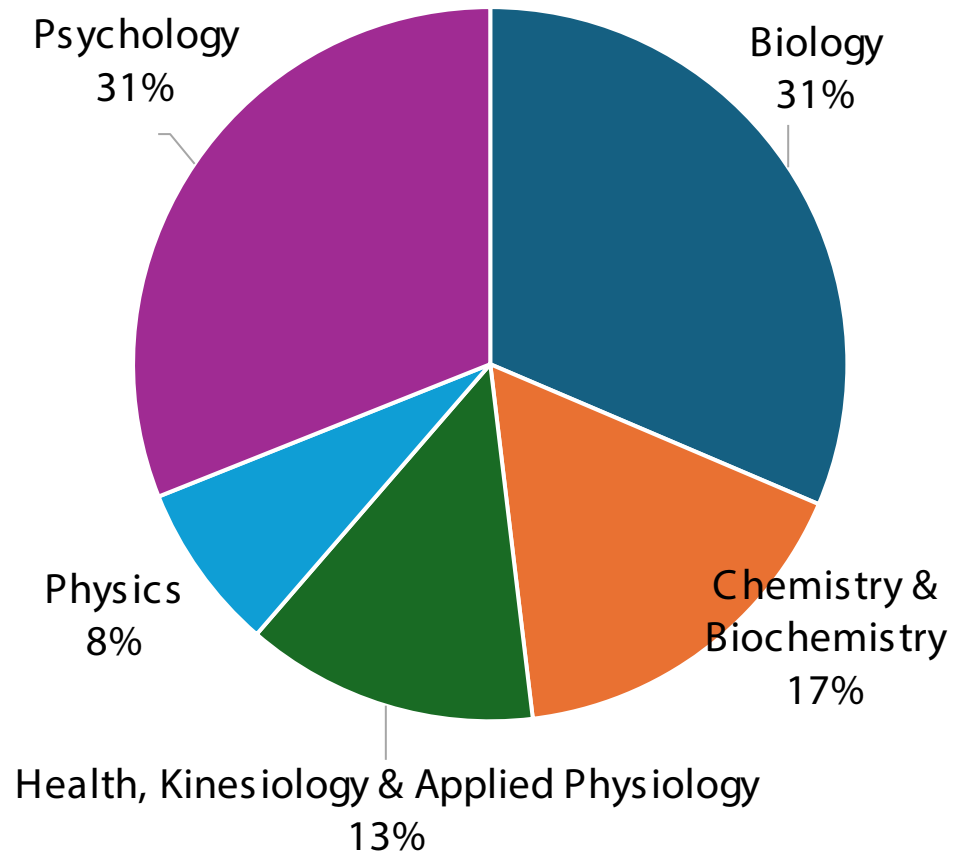
FOCUS: Citations that we can confidently say are to a preprint

We identified 156 citations to preprints in 92 theses from a sample of 815.

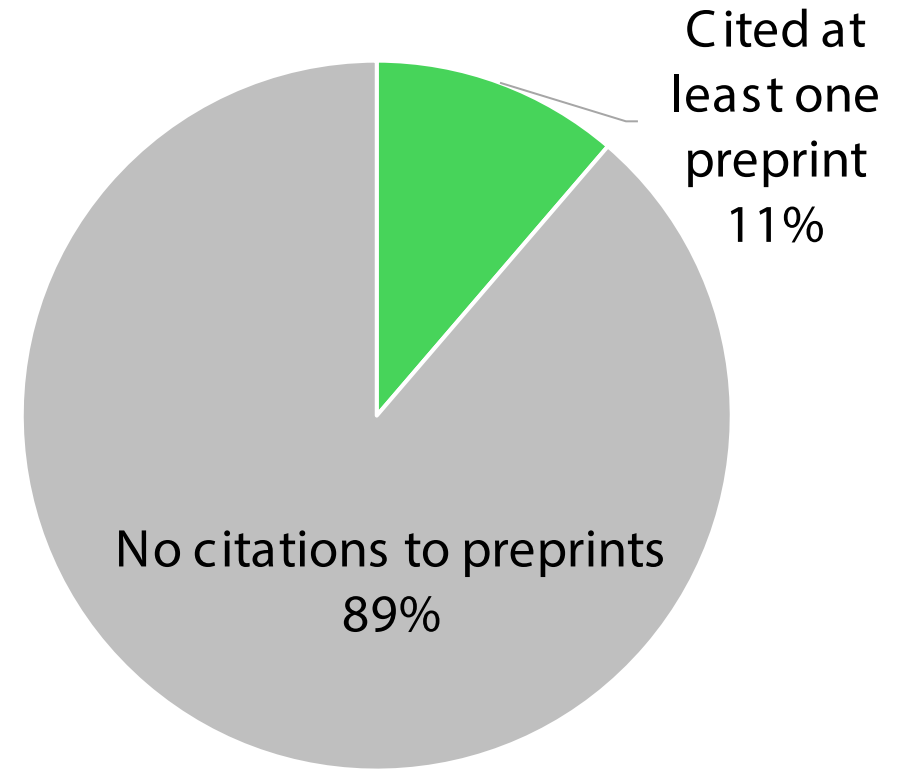
We have 134 challenging citations awaiting final decision

Preliminary Results: Our Sample (n=815)

Total theses & dissertations

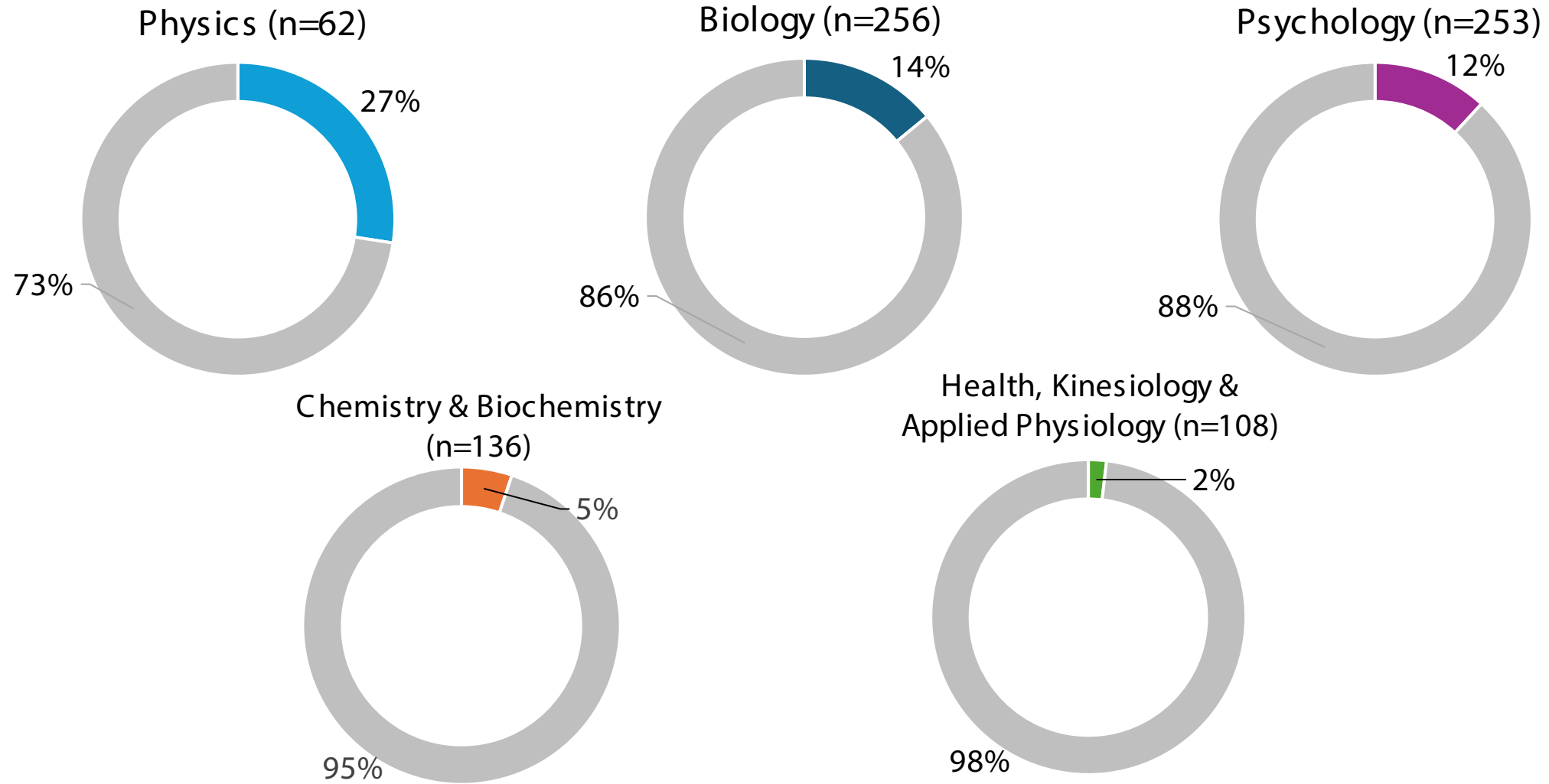


Theses with preprint citations



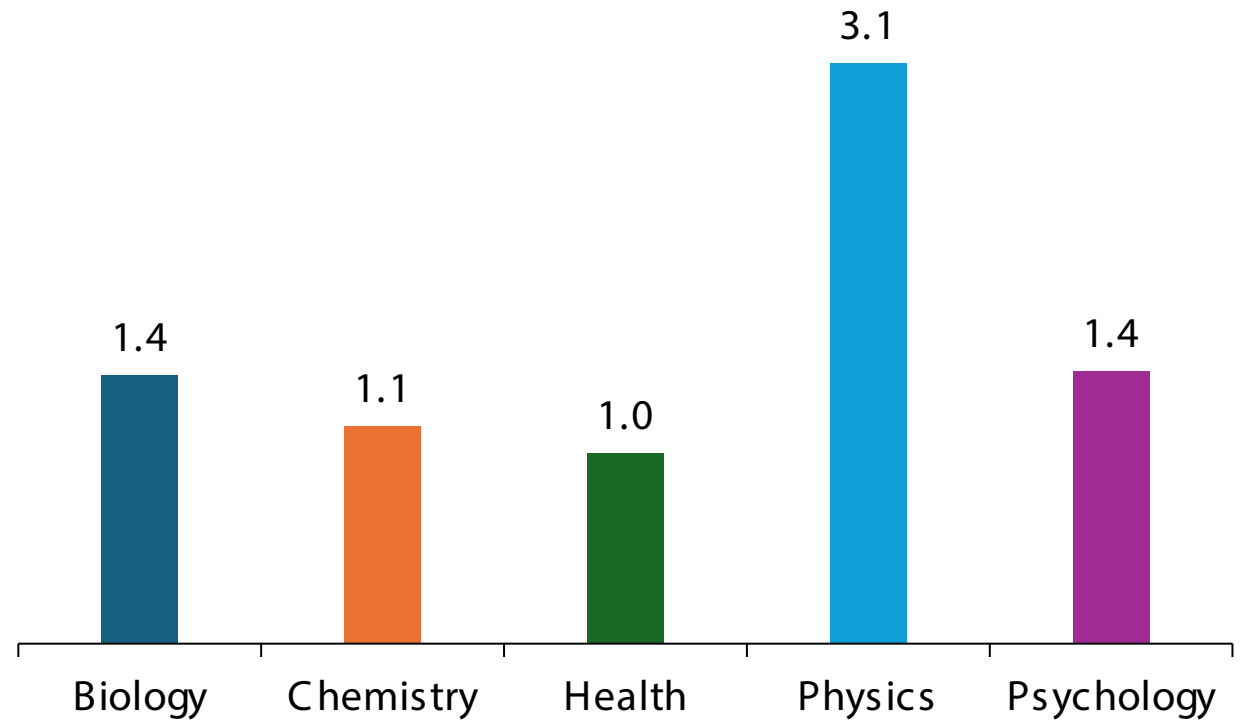
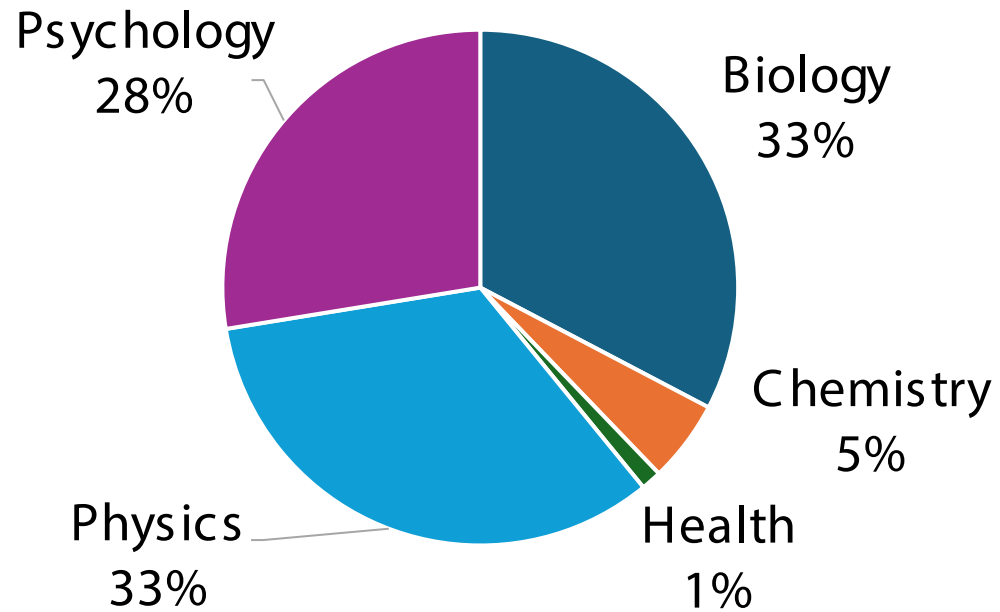
Preliminary Results:

Theses with preprint citations, by discipline



Preliminary Results: Citations to preprints by discipline

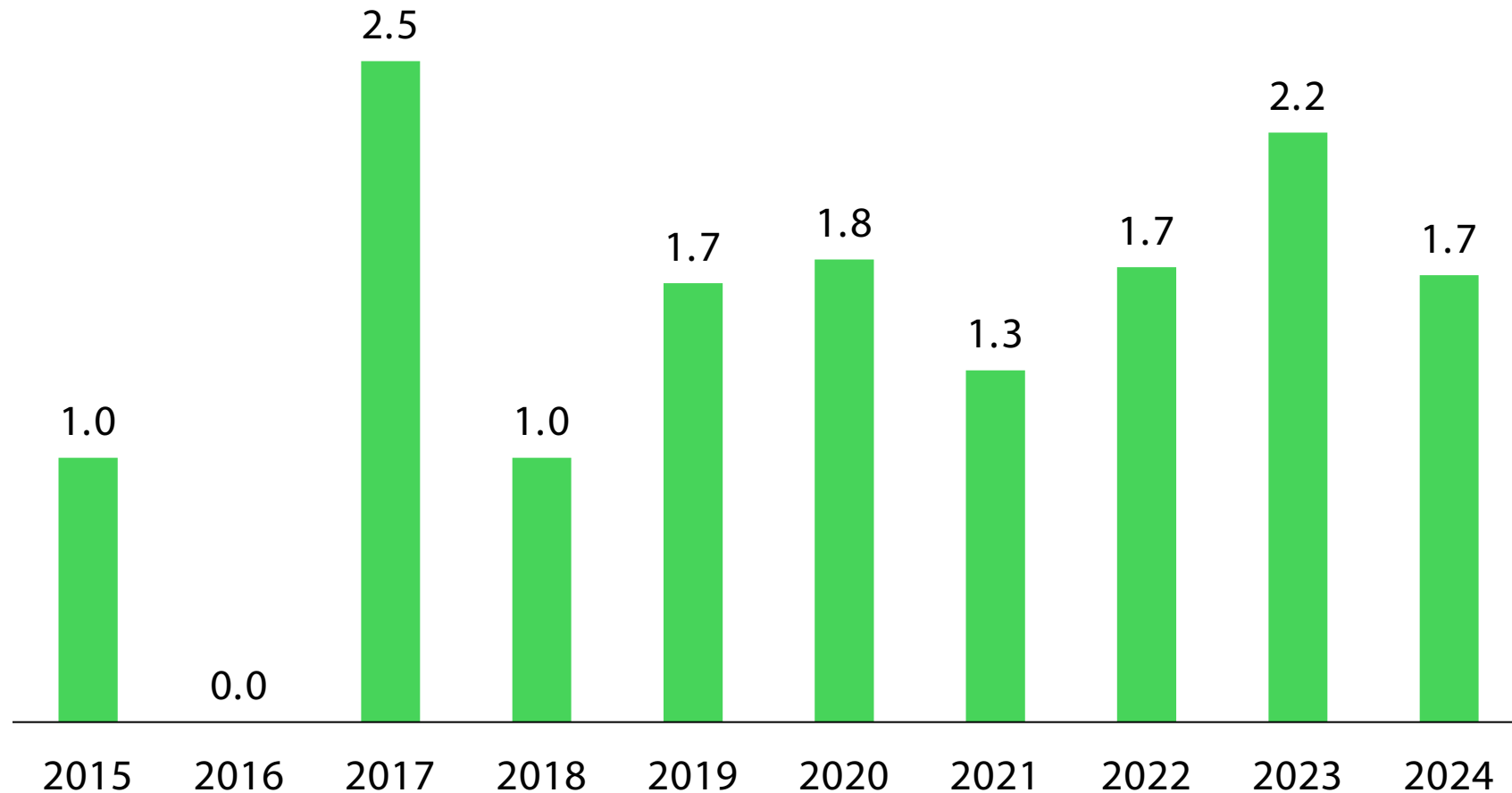
Total preprint citations (n=156)



Average preprint citations per thesis

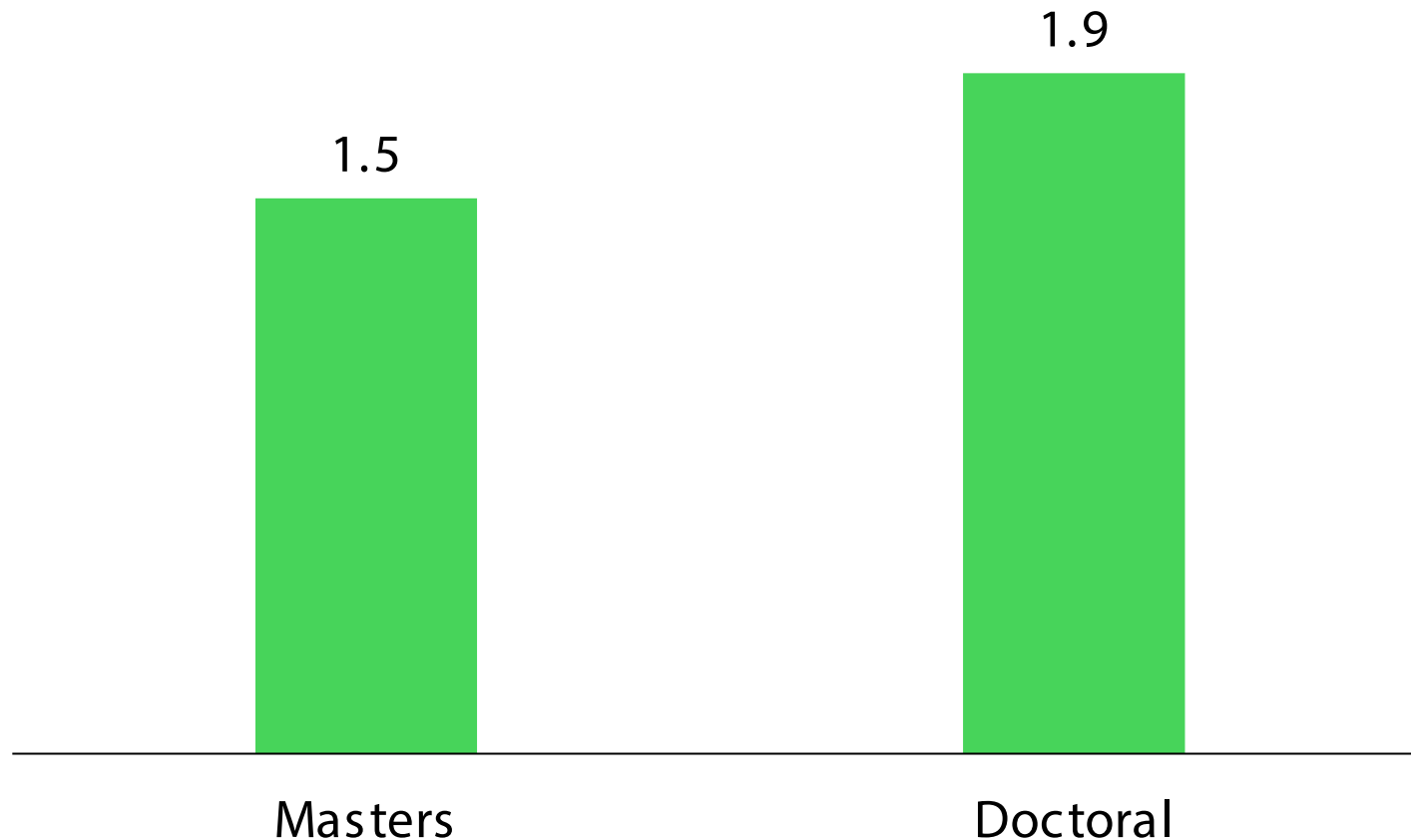
Preliminary Results:

Average preprint citations per thesis over time

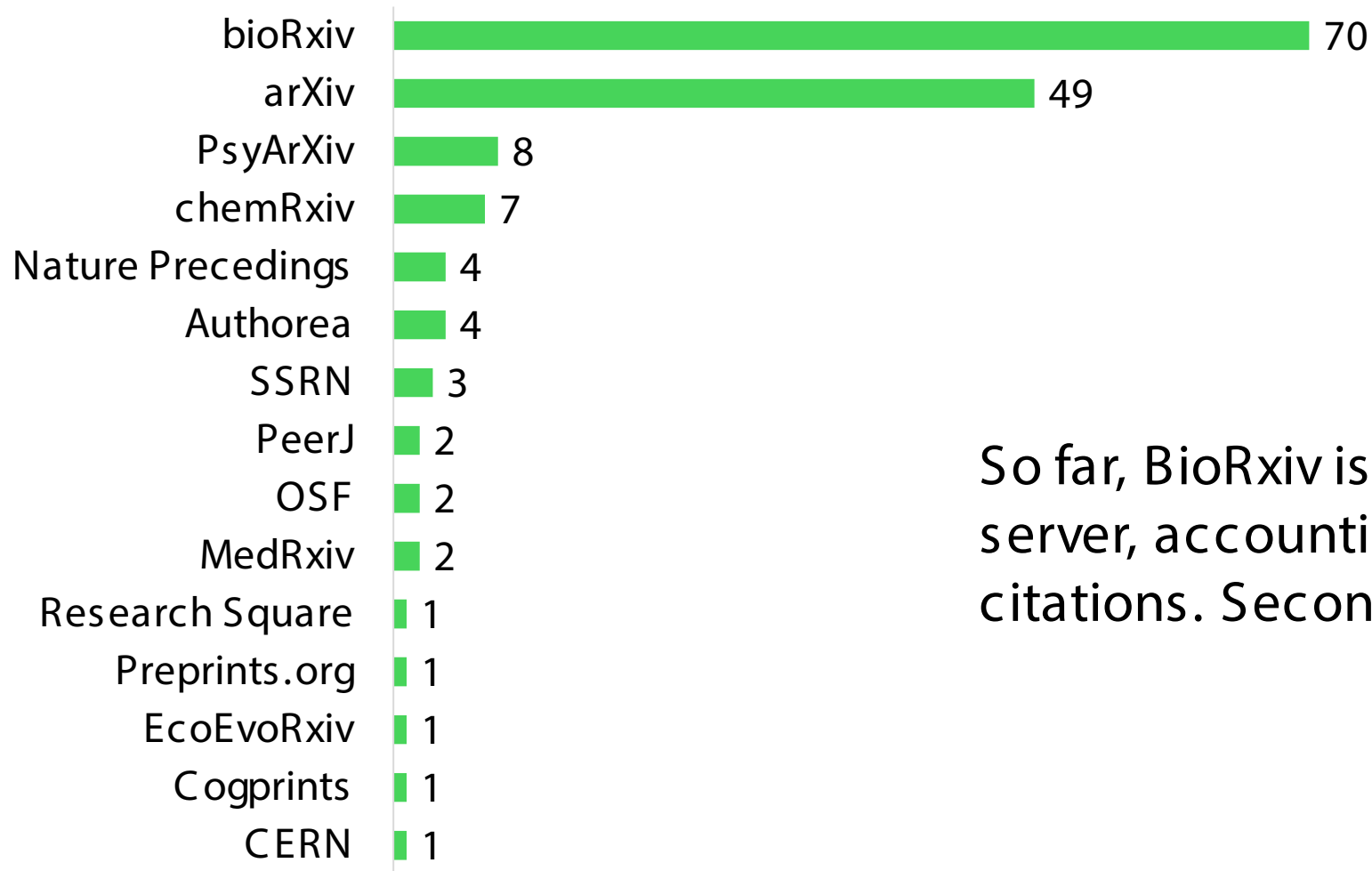


Preliminary Results:

Average preprint citations per thesis by level



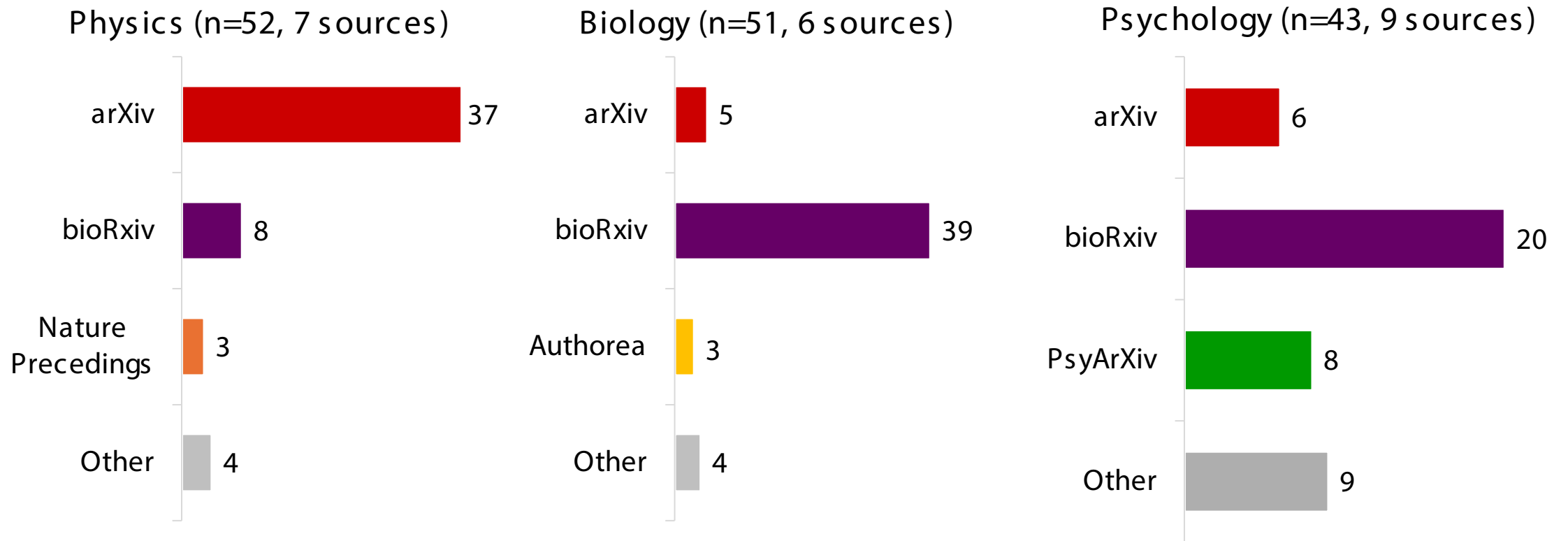
Preliminary Results: Preprint servers cited



So far, BioRxiv is the top cited preprint server, accounting for 45% of all preprint citations. Second is ArXiv with 31%

Preliminary Results:

Preprint servers cited, by discipline



Still to come...

- Decisions on challenging cases
- Further data analysis
- Manuscript



Thank you! Questions?