

Changing the shoulders I am standing on:

Describing the changes that occurred in publications' reference lists after peer review

Aliakbar Akbaritabar, Dimity Stephen, Judith Hartstein & Christophe Heger

PEERE Conference, 30 September 2020

akbaritabar | stephen@dzhw.eu

Introduction

Recent studies in the social sciences (1,2) found that peer review prompted **changes predominantly in the theoretical framing of a study**, and substantially less changes in the methodology.

Reference lists changed accordingly, **increasing by on average 26%** in the published version, although the degree of change was correlated with the extent of theory-based criticism.

References were **not just added, but some removed and new ones** added to align with the change in framing.

1. Strang, D. & Siler, K. (2015). Revising as reframing: Original submissions versus published papers in *Administrative Science Quarterly*, 2005 to 2009. *Sociological Theory*, 3(1), 71-96. DOI: 10.1177/0735275115572152.
2. Teplitskiy, M. (2016). Frame search and re-search: How quantitative sociological articles change during peer review. *The American Sociologist*, 47(2-3), 264-288. DOI: 10.1007/s12108-015-9288-3.

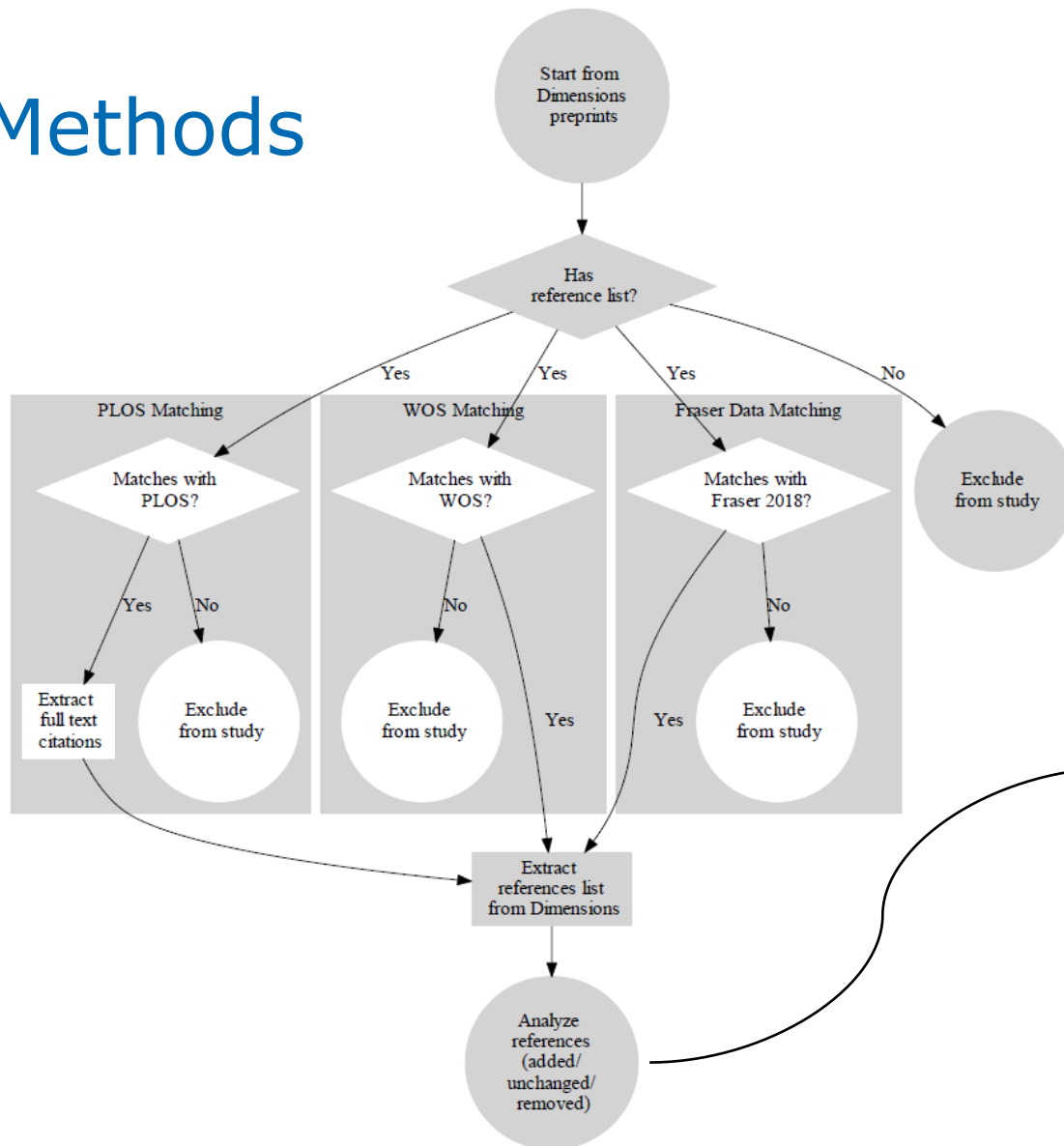
Research questions

Do these findings from the social sciences also apply to other fields?

We will match preprints and publications and compare their references lists to see

- i) how their reference lists changed, and
- ii) which sections of manuscripts underwent the most change.

Methods



Peer Review Effect
Data Gathering and Match Process

Preprint	Publication	Status
Ref. 1	Ref. 1.	Unchanged
Ref. 2	Ref. 2	Unchanged
Ref. 3	--	Removed
Ref. 4	Ref. 4	Unchanged
Ref. 5	--	Removed
--	Ref. 6	Added

Status	Count	Proportion
Unchanged	3	0,50
Removed	2	0,33
Added	1	0,17

Results

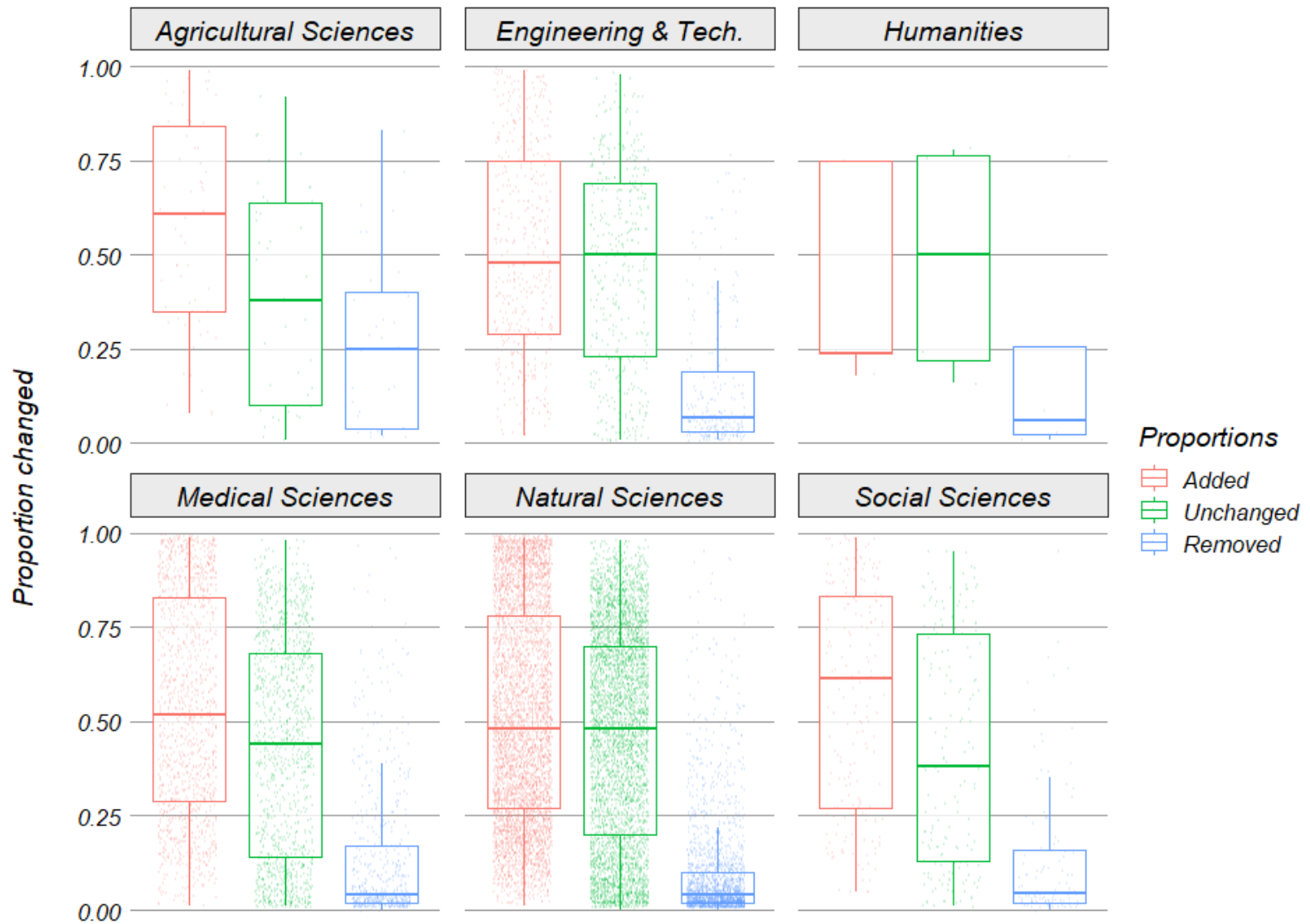
Source	No. pairs*
Fraser	3,250
WOS	2,939
PLOS	557
Total	6,746

Fields	No. pairs**
Natural sciences	4,903
Medical sciences	1,344
Engineering & tech.	460
Social sciences	201
Agricultural sciences	59
Humanities	5

Year	No. preprints	No. publications
2002	1	0
2003	0	1
2004	0	0
2005	0	0
2006	0	0
2007	3	2
2008	0	1
2009	1	1
2010	9	6
2011	4	7
2012	1	0
2013	14	11
2014	210	142
2015	22	80
2016	86	13
2017	3,965	1,127
2018	2,430	5,345
2019	na	10

* includes double-counting of pairs identified in multiple sources, ** or assigned to multiple disciplines

Proportions of references added, unchanged or removed by field



Results

Kruskal-Wallis test results

Test	χ^2	df	p	effect size
Added ~ Field	15.266	3	< 0.01	0.002
Removed ~ Field	64.083	3	< 0.001	0.01
Unchanged ~ Field	15.614	3	< 0.01	0.002

Dunn tests for differences between fields

Level	Fields	z	P (adj.)
Added	Medical sciences v Natural sciences	-3.20	< 0.01
Removed	Engineering v Medical sciences	-3.13	0.01
Removed	Engineering v Natural sciences	-6.87	< 0.001
Removed	Medical sciences v Natural sciences	-4.81	< 0.001
Unchanged	Medical sciences v Natural sciences	3.65	< 0.01

Part II: Section changes

We used hyperlinked full-text from PLOS to identify which sections of the text the references were in.

Introduction

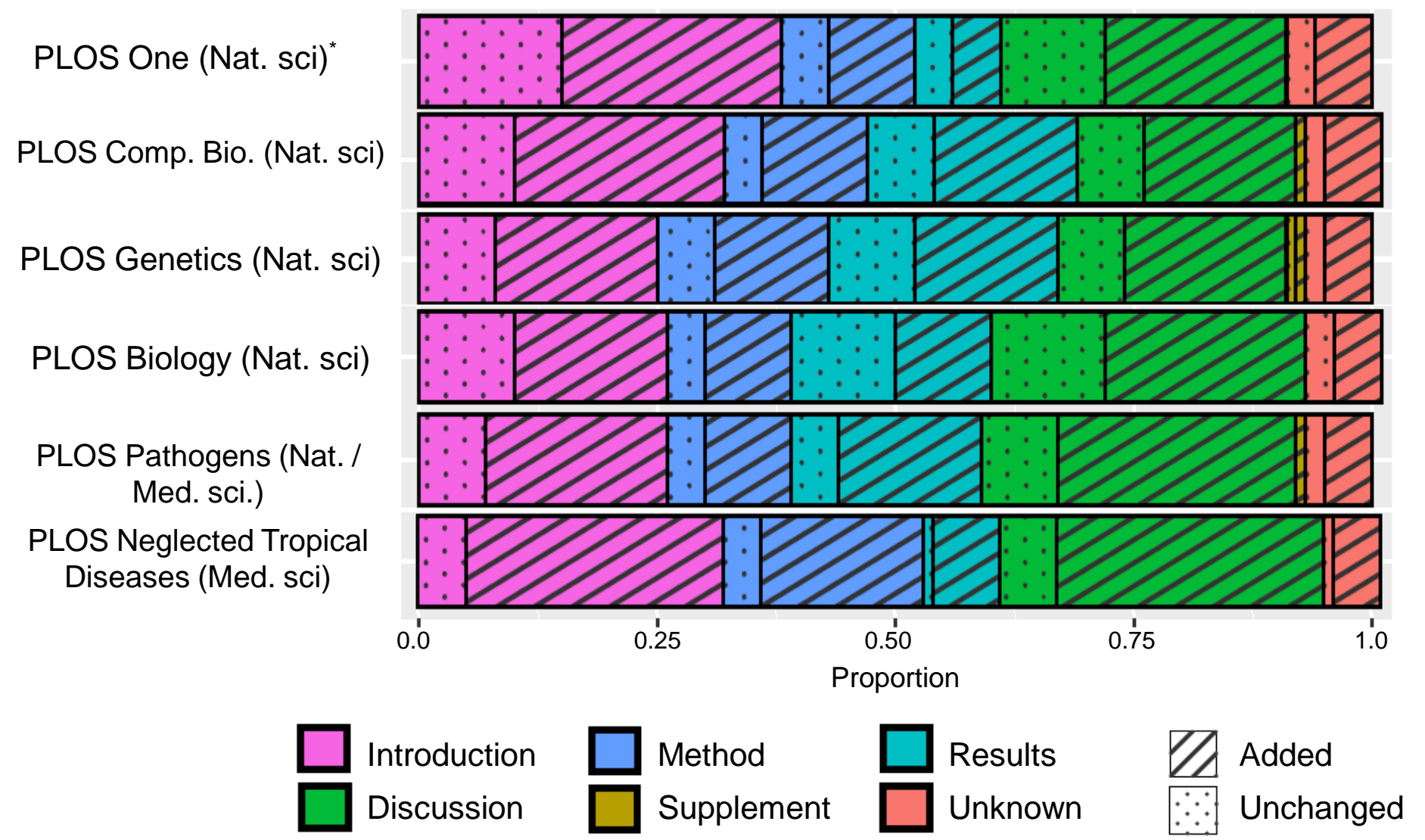
In eukaryotes, the MADS-box gene family encodes transcription factors that play important roles in numerous biological functions by encoding transcription factors [1]. MADS-box transcription factors are characterized by the presence of a DNA binding domain that is approximately 60 amino acids length, known as the MADS domain, located at the N-terminal region of the protein. In plants, one of the most significant features of the MADS-box gene family is its essential role in the ABCDE model of flowering [2]. Numerous studies have identified its vital function in the formation and growth of floral organs [3], anthesis time [4], ovule development [5] and the ripening of fruits and seeds [6]. MADS-box gene family members have also been reported to be involved in stress responses [7], including abiotic and biotic responses [8]. For instance, the expression of *TaMADS2* was up-regulated after being infected by stripe rust fungus in wheat [9], and some MADS-box genes may also be involved in response to high salt concentrations [8]. In addition, the MADS-box gene plays an important role in the development of roots and trichomes [10].

In animals, plants and fungi, the MADS-box gene family has been categorized into two main groups, i.e., type I and type II (MIKC) [10,11]. Type I has been further classified into M-type and N-type genes [12]. In addition to the MADS domain, type II genes contain three additional domains, i.e., an Intervening domain, Keratin-like domain and C-terminal domain. Type II genes can be divided into MIKC^C-type and MIKC^N-type genes based on the structural divergence of the Intervening domain [13,14]. A Bayesian classification of the MADS-box proteins in *Arabidopsis* categorized the proteins into five distinct groups (Ma, Mb, My, Ms, and MIKC) [6]. This classification scheme is adopted throughout the present study.

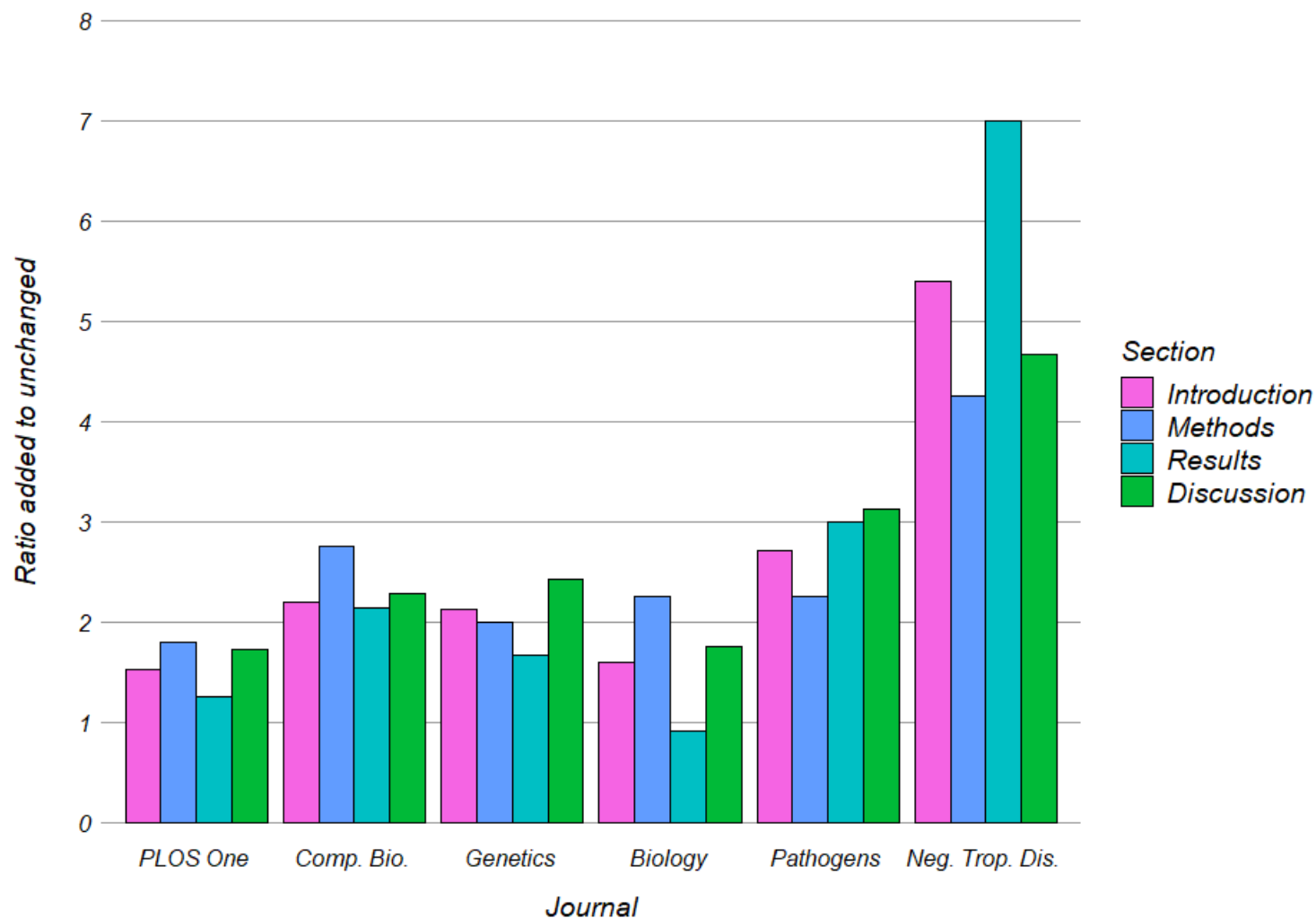
Given its important roles, the MADS-box gene family has been widely characterized in many plant species, including *Arabidopsis* [6], *Brachypodium* [8], rice [15], maize [16], sorghum [16], apple [17], poplar [14], cucumber [18] and soybean [19]. Common wheat (*Triticum aestivum* L., AABBDD genome, 2n = 6x = 42) is one of the most important cereal crops. To date, a genome-wide analysis of MADS-box gene family in wheat is yet to be published. The publication of the draft genome sequences of the common wheat cultivar 'Chinese Spring' [20] and its two

```
<?xml version="1.0"?>
<sec id="sec001" sec-type="intro">
  <title>Introduction</title>
  <p>
    In eukaryotes, the MADS-box gene family encodes transcription factors that play
    important roles in numerous biological functions by encoding transcription factors [
    <xref ref-type="bibr" rid="pone.0181443.ref001">1</xref>
    ]. MADS-box transcription factors are characterized by the presence of a DNA binding
    domain that is approximately 60 amino acids length, known as the MADS domain, located
    at the N-terminal region of the protein. In plants, one of the most significant
    features of the MADS-box gene family is its essential role in the ABCDE model of
    flowering [
    <xref ref-type="bibr" rid="pone.0181443.ref002">2</xref>
    ]. Numerous studies have identified its vital function in the formation and growth of
    floral organs [
    <xref ref-type="bibr" rid="pone.0181443.ref003">3</xref>
    ], anthesis time [
    <xref ref-type="bibr" rid="pone.0181443.ref004">4</xref>
    ], ovule development [
    <xref ref-type="bibr" rid="pone.0181443.ref005">5</xref>
    ] and the ripening of fruits and seeds [
    <xref ref-type="bibr" rid="pone.0181443.ref006">6</xref>
    ]. MADS-box gene family members have also been reported to be involved in stress
    responses [
    <xref ref-type="bibr" rid="pone.0181443.ref007">7</xref>
    ], including abiotic and biotic responses [
    <xref ref-type="bibr" rid="pone.0181443.ref008">8</xref>
    ]. For instance, the expression of
    <italic>TaMADS2</italic>
    was up-regulated after being infected by stripe rust fungus in wheat [
    <xref ref-type="bibr" rid="pone.0181443.ref009">9</xref>
    ], and some MADS-box genes may also be involved in response to high salt concentrations
    [
    <xref ref-type="bibr" rid="pone.0181443.ref008">8</xref>
    ]. In addition, the MADS-box gene plays an important role in the development of roots
    and trichomes [
    <xref ref-type="bibr" rid="pone.0181443.ref010">10</xref>
    ].
  </p>
</sec>
```


Mean proportion of references added or unchanged by journal section



Ratio of proportion of added to unchanged references by section and journal



Discussion

We found that:

Manuscripts across all fields undergo extensive bibliographic changes

1. $\geq 50\%$ of references are added in publication
2. Removing references is much less common

Medical and natural sciences differed on all levels

1. Medical sciences added and removed a larger proportion
2. Natural sciences referencing was more stable

A larger proportion of references were added to introduction / discussion, but largest relative change was in method and results sections

Conclusions & limitations

There are **significant differences** between fields in how referencing changes.

The **primary focus** of social sciences on theoretical framing may not extend to the medical sciences or natural sciences.

We had a much **larger sample size** ($n = 6,002$) than other studies ($n = \sim 70$).

However!

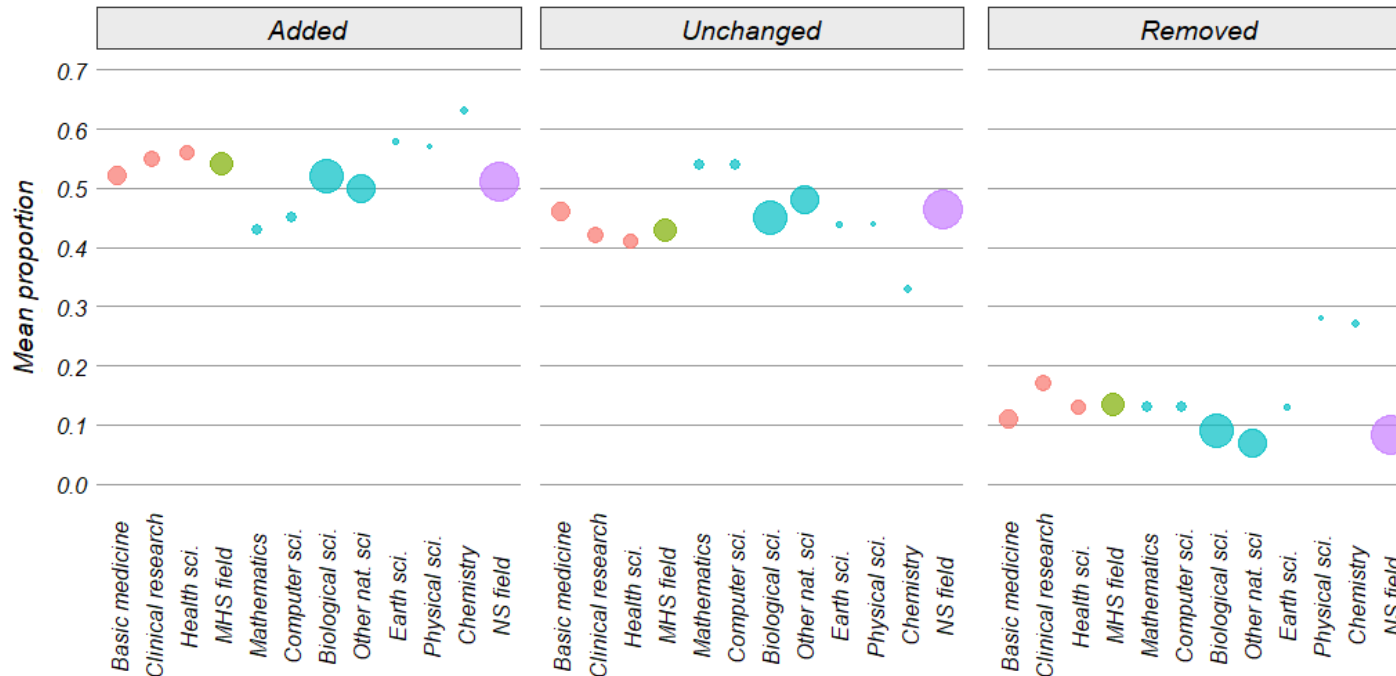
- **Preprint \neq manuscript** ready for submission
- Changes are **not purely due to peer review**

Limitations:

- Reference lists in Dimensions are not always complete
- Dimensions' internal matching is not perfect
- We couldn't identify the sections of 6-9% of references, but do know if they were added or unchanged

Next steps

Can discipline-level analysis explain more variability in reference changes?



Do multidisciplinary papers change more than „pure“ papers to align them with disciplinary traditions?

Does the focus of peer review change by journal factors, e.g. impact, OA status?

Can we detect patterns of suggested citations to the publishing journal?

Thank you!

Please get in touch if you have any questions or comments:

akbaritabar@dzhw.eu / stephen@dzhw.eu

We thank Nick Fraser, Fakhri Momeni, Philipp Mayr, and Isabella Peters, who kindly allowed us to use their data to complement our own dataset.

Data were obtained via the German Competence Center for Bibliometrics, which is funded by the German Federal Ministry for Education and Research (BMBF), grant number 01PQ17001.