

# Are Successful Co-authors More Important than First Authors for Publishing Academic Journal Articles?<sup>1</sup>

Mike Thelwall, University of Wolverhampton

Academic research often involves teams of experts, and it seems reasonable to believe that successful main authors or co-authors would tend to help produce better research. This article investigates an aspect of this across science with an indirect method: the extent to which the publishing record of an article's authors associates with the citation impact of the publishing journal (as a proxy for the quality of the article). The data is based on author career publishing evidence for journal articles 2014-20 and the journals of articles published in 2017. At the Scopus broad field level, international correlations and country-specific regressions for five English-speaking nations (Australia, Ireland, New Zealand, UK and USA) suggest that first author citation impact is more important than co-author citation impact, but co-author productivity is more important than first author productivity. Moreover, author citation impact is more important than author productivity. There are disciplinary differences in the results, with first author productivity surprisingly tending to be a disadvantage in the physical sciences and life sciences, at least in the sense of associating with lower impact journals. The results are limited by the regressions only including domestic research and a lack of evidence-based cause-and-effect explanations. Nevertheless, the data suggests that impactful team members are more important than productive team members, and that whilst an impactful first author is a science-wide advantage, an experienced first author is often not.

**Keywords:** journal impact; collaboration; co-authorship; citation counts; journal impact factors

## Introduction

In many fields, articles are the result of a long research process, with the publishing journal serving quality assurance, dissemination, and reputation building functions. High reputation journals are particularly desirable for this because they often have the strictest quality control, attract the widest audience, and confer the most reputational capital on the authors. Whilst Journal Impact Factors (JIFs) are frequently misused as journal or article quality measures, this is inappropriate (Kurmis, 2003; Seglen, 1998). Nevertheless, higher impact journals are statistically more likely to be high quality (Mahmood, 2017) and articles in higher impact journals are statistically more likely to be influential or otherwise high quality (e.g., Waltman & Traag, 2020). Investigating factors that statistically associate with publishing in higher impact journals may therefore help researchers to design more influential or higher quality research.

Putting together a team and (usually) attracting funding are common first steps in research. This article focuses on the relative roles of the first author and their team in getting published in higher impact journals, as a proxy for conducting high quality research. It uses a quantitative approach and a set of simplifications and assumptions to allow a science-wide analysis. The first author of an article tends to be the main contributor in most fields, despite a degree of alphabetical ordering in some, senior researchers being listed last in others (Larivière et al., 2016) and corresponding authors also sometimes being important (Grácio et

---

<sup>1</sup> Thelwall, M. (in press). Are successful co-authors more important than first authors for publishing academic journal articles? *Scientometrics*

al., 2020). A new or less successful author may seek a more experienced author as a collaborator in the belief that their work could be improved, such as through access to more powerful equipment, funding, better disciplinary knowledge, or help with publishing (e.g., Amjad & Munir, 2021; Uwizeye et al., 2020; van Rijnsoever & Hessels, 2021). The value of this common sense strategy does not seem to have been directly tested before, however, except for mentor-mentee relationships. Identifying the relative importance of first and experienced or successful authors may therefore help scholars to decide whether seeking collaborators may be useful.

Team science is particularly important for tackling multidisciplinary problems and bringing together complementary sets of expertise (Gibbons, et al., 1994). Many studies have investigated the role of teams in the production of science from heterogeneous perspectives (Hall et al., 2018), as briefly summarised here. The number of co-authors per paper increases slightly with JIFs in some fields but not others (Abramo & D'Angelo, 2015; Glanzel, 2002). Articles with more authors tend to be more cited (Katz & Hicks, 1997), but this varies between fields (Abramo & D'Angelo, 2015), countries (Thelwall & Maflahi, 2020), and over time (Larivière, et al., 2016). Co-authorship benefits seem likely to differ substantially between collaboration types, such as peer collaborations compared to mentor-mentee collaborations (Larivière, 2012), for large-scale collaborations (Thelwall, 2020a) and for long-term collaborations (Bu et al., 2018a). Internationally co-authored articles also tend to be more cited, despite not being more novel (Wagner et al., 2019). In Italy, at least, author success associates with the ability to collaborate internationally (Abramo et al., 2019). The relationship between author properties for different positions in the authorship list has also been investigated, but mainly for gender effects (Thelwall, 2020a). Nevertheless, collaborating with a successful scholar in some fields can improve career publishing outcomes (Bu et al., 2018b), even if not productivity (Levitt & Thelwall, 2016).

The relative importance of journals varies greatly between fields for structural reasons. Some specialities may have a group of journals with similar quality reputations that primarily segment by topic (e.g., *Fashion Theory: Journal of Dress Body and Culture* vs. *Rock Art Research*). Other fields may instead tend to have journals that publish similar topics (e.g., *Scientometrics* vs. *Quantitative Science Studies* vs. *Research Evaluation* vs. *Journal of Informetrics* vs. *Journal of Scientometric Research*), where practitioners may believe that some of them tend to publish more important work or have higher quality control. Most fields are probably hybrids, with a combination of partly specialised but overlapping journals and recognised generalist journals (e.g., Nicolaisen & Frandsen, 2015) that tend to attract the best articles on any topic (e.g., *Science*, *Nature*). The rise of megajournals has complicated the situation by introducing a class of journal that is generalist but may have variable quality control mechanisms due to the size or fluctuating nature (for special issues) of the editorial team, or that may disregard some traditional quality criteria (Wakeling et al., 2019). Nevertheless, in some fields, recognised hierarchies of journal importance may be fairly well reflected by JIFs (Kelly et al., 2013; Mahmood, 2017; Steward & Lewis, 2010), and so the citation rate of a journal in those fields could be a reasonable indicator of its quality relative to other journals in the same field. This, in turn, may be a reasonable indicator of the quality of the individual articles in that journal given that the quality of a journal is equal to the average quality of its articles. In fact, it may be a better indicator of the quality of an article than the number of citations received by that article (Waltman & Traag, 2020). In fields where JIFs are often regarded as an end in themselves, equating journal quality with JIFs may also give a positive feedback loop as higher-JIF journals can attract more papers and can become

more competitive to publish in. This also applies to researchers in countries that use JIFs to help construct lists of recommended journals (Pölonen et al., 2021).

This article is motivated by assessing the extent to which the publishing record of the first author and their team might influence the quality of their article. For practical reasons, it does not investigate this directly but instead focuses on the statistically related (and measurable) issue of the extent to which the publishing record of the first author and their team might influence their article being published in a more cited journal. Whilst more cited journals are not necessarily more influential, even within a field, they tend to be in some fields, as argued above. The focus is on the first author, who tends to be the main contributor in most fields (Larivière et al., 2016; Mongeon, et al., 2017). Even though alphabetical ordering is common, but not universal, in some fields (Levitt & Thelwall, 2013), the first author seems more likely to be the most important author in these because of the non-alphabetically ordered share of articles. The most successful author also seems relevant for multi-field comparisons, since the senior author might be second in some fields and last in others (e.g., Mongeon, et al., 2017), and may be less influential than a successful more junior author. The corresponding author is also important or the most senior author in some fields (Grácio et al., 2020), but it is primarily an administrative role, not a reliable indicator of seniority (da Silva et al., 2013; Ding & Herbert, 2022), and its purpose seems to vary substantially between countries (Mattsson et al., 2011) and fields (Yu & Yin, 2021). Since the corresponding author is mostly likely to be the first author (Mattsson et al., 2011) and the most important author is also most likely to be first, with field variations, the focus here is only on the first author as the most consistently important author. Co-authorship is investigated although there are many other types of collaboration (Katz & Martin, 1997; Laudel, 2002). The following research questions drive the quantitative part of this study.

- RQ1: Is an article more likely to be published in a higher impact journal if its first author publishes more articles?
- RQ2: Is an article more likely to be published in a higher impact journal if any of its authors publish more articles?
- RQ3: Is an article more likely to be published in a higher impact journal if its first author publishes more cited articles?
- RQ4: Is an article more likely to be published in a higher impact journal if any of its authors publish more cited articles?
- RQ5: Which is more influential, the first author or the team maximum in RQs 1-4?

## Methods

The overall research design was to obtain a large sample of articles over an extended period for all of science and then to derive indicators to statistically assess the association between journal impact and author properties for one of the years. The bibliometric database Scopus was chosen as the source of journal articles because it has science-wide coverage and seems to include more articles than the Web of Science (Martín-Martín, et al., 2021), and with a slightly finer-grained field structure (Scopus, 2020), which helps with the field normalisation steps of the citation analysis component.

The years 2014–2020 were chosen as the period of analysis. This gives seven years, which seems reasonable to assess the productivity of the authors. Whilst a longer period (e.g., 40 years) would be closer to the full publishing lifespan of the authors involved, early publishing records may not be relevant to later publishing and so a narrower window seems more appropriate. Seven years was chosen to give three years either side of the year to be

analysed, 2017, so that all publishing records examined for authors are within three years of each article analysed, keeping them relatively close in time. Author publishing activities were therefore evaluated over seven years, 2014-20.

The middle year, 2017, was chosen as the year to examine individual articles and their publishing journals. Although the final year would be more natural because it would take into account the achievements of the team before they had published each article analysed, a middle year instead focuses on the average achievements and capability of the team at around the time when the article examined was published. This is reasonable since research capability does not seem to grow substantially during careers, at least as reflected in bibliometric indicators (Maflahi & Thelwall, 2021). The year 2017 also gave a citation window of three years, which is sufficient for most fields (Wang, 2013).

Since the primary goal of this study is to compare first author and team contributions, solo articles are in the unique position of having the same first author and team. Whilst these articles could reasonably have been excluded for all analyses, they were kept so that the data for the analyses is more comprehensive.

### *Data*

The data analysed in this article consists of all Scopus-indexed journal articles from 2014-2020, as downloaded in January 2021 using the Scopus API. Scopus organises journal articles into 27 broad fields, each of which contains one or more narrow fields. There are 334 narrow fields in use, but not all years contain articles in all fields. An article is allocated to a set of narrow fields based on its publishing journal. Journals are typically assigned multiple narrow fields to cover all aspects of its scope. For example, *Scientometrics* is currently in three narrow fields: General Social Sciences [Social Sciences]; Library and Information Sciences [Social Sciences]; Computer Science Applications [Computer Science].

All articles were counted at full value within narrow fields for field normalisation purposes, but duplicates were eliminated when narrow fields were merged into broad fields. For example, each *Scientometrics* article was a single (whole) data point for each analysis (regression or correlation) for both the Computer Science and Social Sciences broad fields.

### *Indicators*

The impact of each journal was assessed by calculating its Mean Normalised Log-transformed Citation Score (MNLCS) for the year examined, 2017. The MNLCS (Thelwall, 2017) is a field normalised indicator based on averaging the log-transformed citation counts  $\ln(1+X)$  for all articles in the journal and then dividing by the average of the log-transformed citation counts for all articles in the field of the journal. If the journal was in multiple Scopus narrow fields, then it was instead divided by the average of the field averages. The log transformation greatly reduces the skewing of a typical set of citations and prevents the result from being unduly influenced by a single article. It is therefore less sensitive than the traditional JIF. A unit was added to each citation count before the log transformation,  $\ln(1+x)$ , to avoid errors for uncited articles. The MNLCS for a journal is 1 if the journal's articles are average cited for the field(s) and year examined, above 1 if they tend to be more cited, and below 1 if they tend to be less cited. MNLCS values can be reasonably compared between fields because of this property, although this is not needed here. Note that the journal MNLCS is calculated based on citations to articles in 2017, unlike the traditional JIF, which counts citations from a given year to the previous two years.

The average citation impact of each author was calculated in a similar way: their MNLCS is the average NLCS (normalised log-transformed Citation Score) of all articles published by that author (in any authorship position) 2014-20. The NLCS is the log-transformed citation count divided by the average log-transformed citation count of all articles from the same narrow field and year (or the average of these averages if the article is in multiple fields).

There is a small bias in the data since the citations used to calculate author MNLCS values are also used to calculate journal MNLCS values and the relationship between the two is investigated. This level of bias will be very small for most journals since the MNLCS does not allow individual articles to greatly influence values, except for journals publishing few articles.

The publishing productivity of each author was assessed by counting the number of articles in Scopus 2014-20 with the author in the authorship team (whole counting). A second publishing indicator was calculated by counting fractional contributions: if an author was one of  $n$  co-authors then their credit for the article would be  $1/n$  rather than 1. Fractional counting is perhaps a more realistic assessment of an author's contribution to science, but whole counting is more commonly used in scientometrics.

### *Correlations across science*

Spearman correlations were used to assess the relationship between the indicators because some were skewed. Correlation is not able to disentangle associations for overlapping factors, so these results are presented as overall background descriptive information to be followed by more reliable regressions applied to subsets of the data. The correlations were calculated separately for each narrow field but reported aggregated into the 27 Scopus broad fields to show general trends.

### *Regressions for domestic research in single countries*

Regression was used to assess the relative contributions of the different authorship factors to the publishing journal citation rate. A finer-grained approach than for correlation was used to remove spurious factors from the results, focusing on individual countries to avoid the complicating factor of international differences and collaboration. Moreover, only domestic articles were analysed (those with all author affiliations from the same country) to avoid international collaboration effects, which are complex. The regressions were applied to Scopus broad fields rather than the narrow fields used for the correlations to give sufficient data to analyse. Duplicate articles were eliminated first.

To improve the quality of the author publishing information, countries were only analysed if Scopus had relatively comprehensive coverage of the national scientific literature. As of 2015, English was the only language for which the journals were not underrepresented in any major area of science, so this article focuses on large primarily English-speaking nations (USA, UK, Australia, New Zealand, Ireland), to reduce bias due to low Scopus coverage (Mongeon & Paul-Hus, 2016). This excludes many countries with larger English-speaking populations (e.g., India, Pakistan, South Africa, Canada) because a substantial fraction of these countries speaks a different working language (e.g., 57% Hindi, 39% Punjabi, 23% isiZulu, 21% French, respectively). Although English is the first language of only 78% of the USA (with Spanish second at 14%), the USA is particularly well covered by international scholarly databases and so was included.

The dependent variable for the regression was the journal MNLCS and the independent variables were the first author and team MNLCS, the first author and team

publishing productivity, the number of authors of the paper, and the number of institutional affiliations for the authorship team. The last two factors are control variables that are not investigated but are included to at least partially control for authorship teams of different sizes and institutional variety. The inclusion of the number of authors as a control variable is important because larger teams would tend to have higher team statistics, other factors being equal. Also, more collaborative articles tend to be more cited (Abramo & D'Angelo, 2015), giving a second reason to control for team size. Adding institutional affiliations partially controls for a different aspect of the authorship team other than their publication productivity and average impact being an influence.

For the regressions, all variables except journal and author MNLCS were log-normalised with the formula  $\ln(1 + x)$  because of skewing. For consistency, the same log transformation unit offset was used for all variables, even though some had no zero values. The independent variables were normalised by dividing by the standard deviation so that the magnitudes of the regression coefficients (i.e., beta values) could be directly compared. The regression line for 2017 articles from each country is therefore as follows (ignoring the division by the standard deviation).

$$\begin{aligned} \text{MNLCS}_{\text{Journal}} = & \beta_0 + \beta_1 \ln(1 + \text{Authors}) + \beta_2 \ln(1 + \text{Institutions}) \\ & + \beta_3 \ln(1 + \text{Publications}_{\text{FirstAuthor}}) + \beta_4 \ln(1 + \text{Publications}_{\text{MaxAuthor}}) \\ & + \beta_5 \text{MNLCS}_{\text{FirstAuthor}} + \beta_6 \text{MNLCS}_{\text{MaxAuthor}} \end{aligned}$$

To test whether excluding solo articles (where the first and maximum author statistics are the same) influences the results, the regressions were repeated without single author papers. The results were very similar to the main results and are reported only in the online supplement (<https://doi.org/10.6084/m9.figshare.19091687>).

Robust regression (Rousseeuw et al., 2021) was used to minimise the influence of outliers and the Variance Inflation Factor (VIF) was consulted to check for collinearity effects, with values below 5 being accepted (achieved in all cases). The regressions were run for broad fields rather than averaged across narrow fields (as for the correlations) due to the need for sufficient data to get statistically significant regression results. The code and complete results, including VIFs and fitting statistics, are available in the online supplement (<https://doi.org/10.6084/m9.figshare.19091687>).

## Results

Correlations between journal impact (MNLCS) and first author or team research productivity are presented first, followed by correlations between journal impact (MNLCS) and first author or team research impact (MNLCS).

After the correlation results, regression results are given for domestic research for two large English-speaking countries (US and UK; results for three other countries are online), with each regression simultaneously taking into account first author and team productivity and citation impact to show their relative contributions. The regressions are based on broad fields rather than narrow fields. The sample sizes for the analyses are in Table 1.

Table 1. Sample sizes (Scopus-indexed journal articles) for the analyses. The correlations use all articles from 2017 and the regressions use all articles from 2017 from a single country. The 2014-20 data is used to calculate author and team productivity and citation rates. The two “All” columns count by narrow field, so include many duplicate articles occurring in multiple fields. The country columns count by broad field, excluding these duplicates. There are also duplicates between broad fields in all columns, affecting the Total row at the end of the table.

Cluster	Broad field	2017						2014-20
		Australia	Ireland	NZ	UK	USA	All	All
Health Sci	Dentistry	154	15	25	537	1162	13825	100221
Health Sci	Health Professions	1028	111	123	1230	5884	24634	186676
Health Sci	Medicine	10243	1296	1191	16478	92952	626178	4702905
Health Sci	Nursing	1354	160	200	1943	9037	34003	261738
Health Sci	Veterinary	306	46	88	602	2266	19082	140956
Life Sci	Agricultural & Biol Sci	3768	368	814	3308	22541	252596	1830732
Life Sci	Biochem, Gen, Mol Biol	3338	396	469	5811	39547	365240	2779042
Life Sci	Immun & Microbiology	814	121	145	1432	9093	72347	544181
Life Sci	Neuroscience	1204	115	167	2159	12753	67537	495804
Life Sci	Pharmacology	585	86	85	1051	8174	84719	659457
Multi	Multidisciplinary	918	64	132	1686	8087	58089	377435
Physical Sci	Chemical Engineering	992	127	126	1913	9428	121779	911276
Physical Sci	Chemistry	1750	279	198	3486	18956	248061	1848521
Physical Sci	Computer Science	1343	155	210	2776	11671	196609	1513370
Physical Sci	Earth & Planetary Sci	1543	63	272	2113	10812	150017	1078916
Physical Sci	Energy	800	94	73	1428	5634	104247	763319
Physical Sci	Engineering	3179	393	398	5781	23667	468543	3371205
Physical Sci	Environmental Science	2751	213	398	3228	15654	222100	1679077
Physical Sci	Materials Science	1685	241	177	3214	16521	311501	2254507
Physical Sci	Mathematics	983	119	146	2627	12012	133843	982305
Physical Sci	Physics & Astronomy	1580	220	189	4505	21342	275324	1978850
Soc & Hum	Arts & Humanities	2341	336	415	6432	15094	74139	555920
Soc & Hum	Business, Man & Acc	1496	121	255	2312	7660	70989	542682
Soc & Hum	Decision Sciences	323	25	46	663	2912	20813	155217
Soc & Hum	Economic/metrics & Fin	955	136	146	1771	6169	42955	321302
Soc & Hum	Psychology	1955	168	270	3290	16784	65430	490727
Soc & Hum	Social Sciences	6432	740	1055	12441	40921	222656	1696052
<b>All science</b>	<b>Total</b>	<b>53820</b>	<b>6208</b>	<b>7813</b>	<b>94217</b>	<b>446733</b>	<b>4347256</b>	<b>32222393</b>

### *Associations between first author or team productivity and journal citation impact*

Correlations between first and maximum author productivity (number of papers) and journal impact (MNLCS) were calculated in all narrow fields and averaged across broad fields to help identify trends in the results.

### **Correlations between first author productivity and journal citation impact**

In all 27 broad Scopus fields, first authors that wrote more papers 2014-20 tended to get their 2017 paper(s) in slightly more cited journals (i.e., higher 2017 MNLCS) (Figure 1). This is because the average (across narrow fields) Spearman correlation between journal MNLCS and author papers is positive but small (0.05-0.3) in all broad fields. The correlation is always

weaker if fractional counting is used, although the average correlation is still positive in all broad fields except Multidisciplinary. Thus, with this partial exception, *more productive first authors tend to get their work in higher cited journals.*

Within the four disciplinary clusters (Social Sciences & Humanities; Physical Sciences; Life Sciences; Health Sciences) the differences between broad field correlations are large, and there isn't a pattern for any grouping to have substantially larger or smaller correlations than the others. For example, at least one average correlation in each of the four clusters is within the range of values of the other three clusters. Thus, field-specific factors are more important than broad trends in the relationship between first author productivity (number of journal articles) and journal impact (MNLCS). Thus, *whilst there are substantial disciplinary differences in the extent to which more productive first authors tend to get their work in higher cited journals, these do not seem to be systematic in terms of universally higher or lower scores for any of the four disciplinary clusters.*

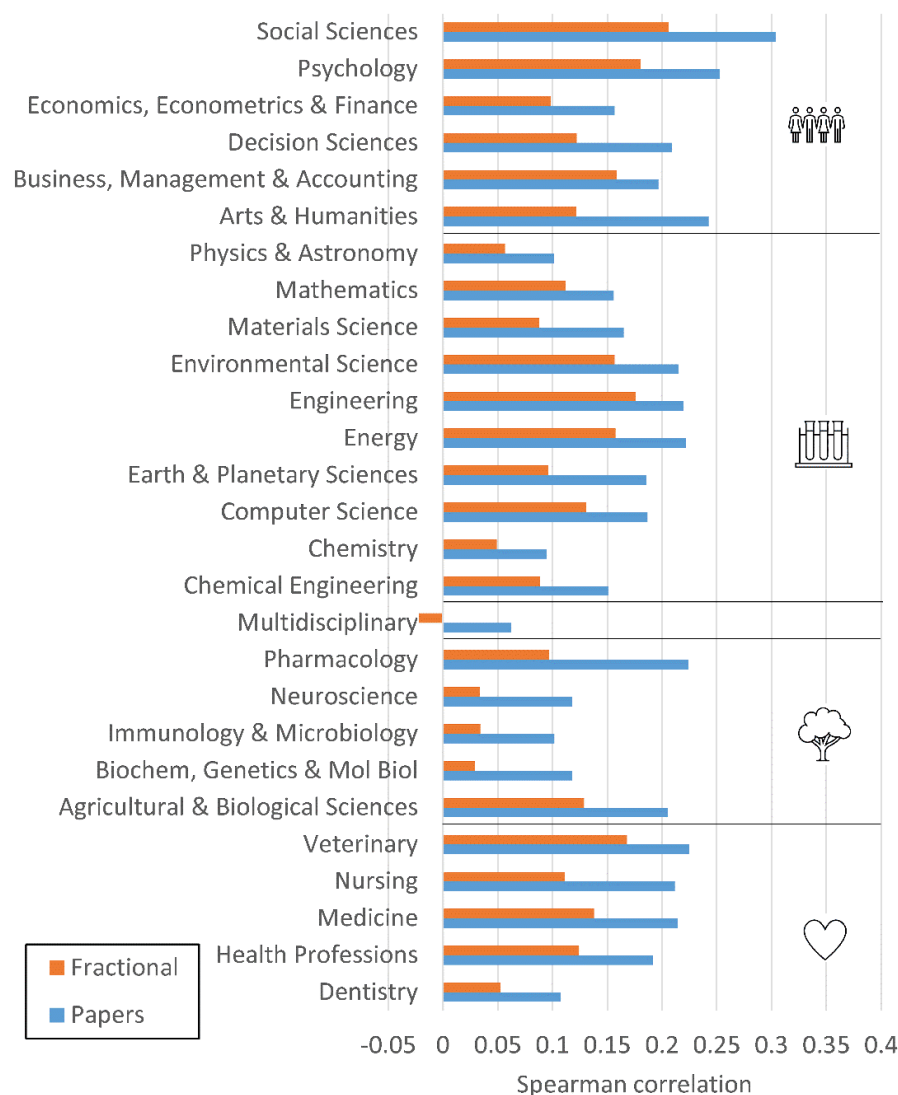


Figure 1. Average (over Scopus narrow fields) Spearman correlations between 2017 journal MNLCS and the fractional or total number of articles 2014-2020 by **the first author**. Broad fields are grouped into the four Scopus clusters, with icons as reminders: Social Sciences & Humanities; Physical Sciences; Life Sciences; Health Sciences. Sample sizes for all graphs are in Table 1.



## Correlations between author team productivity and journal citation impact

In all 27 broad Scopus fields, authorship teams with any author that wrote more papers 2014-2020 tended to get their 2017 paper(s) in more cited journals (i.e., with a higher journal MNLCS) (Figure 2). As above, the correlation is stronger for full than for fractional counting. Nevertheless, the association is only moderate (0.15 to 0.35). Thus, *more productive teams tend to get their work in higher cited journals*.

Since the correlation is strongest for full counting, this is used instead of fractional counting in the remainder of the graphs and for the regressions. Corresponding fractional data is in the supplementary information.

The four disciplinary clusters again overlap in values and are quite similar, so *field-specific factors are more important than broad trends in the relationship between team maximum productivity (number of journal articles) and journal impact (MNLCS)*.

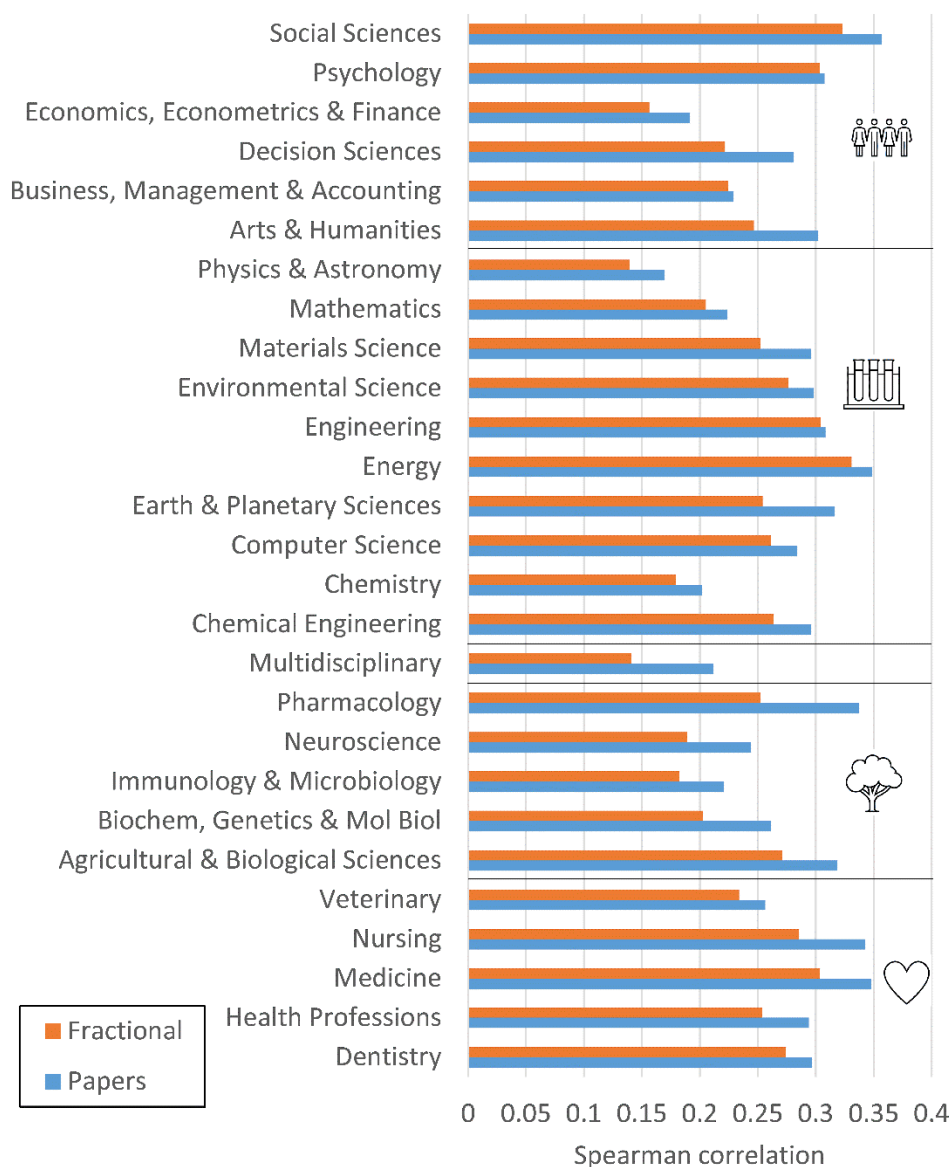


Figure 2. Average (over Scopus narrow fields) Spearman correlations between 2017 journal MNLCS and the fractional or total number of articles 2014-2020 by **the authorship team maximum productivity (articles 2014-20)**.

## First author productivity and author team productivity vs. journal citation impact

Directly comparing the publishing record (number of articles) for the first author with that of the authorship team maximum (Figure 3, combining the relevant Figure 1 and Figure 2 data), suggests that *the authorship team productivity is more important for publishing journal impact than first author productivity in all broad fields.*

Since correlation is not causation, this suggests, but does not prove, that having a higher publishing member of an authorship team is better for getting an article published in a higher impact journal than having a higher publishing first author.

The difference between the first and maximum author correlations is generally smaller in the Social Sciences & Humanities cluster than in the other three clusters (Figure 3). This is presumably a side effect of smaller team sizes in these areas and the commonness of solo research in the arts and humanities.

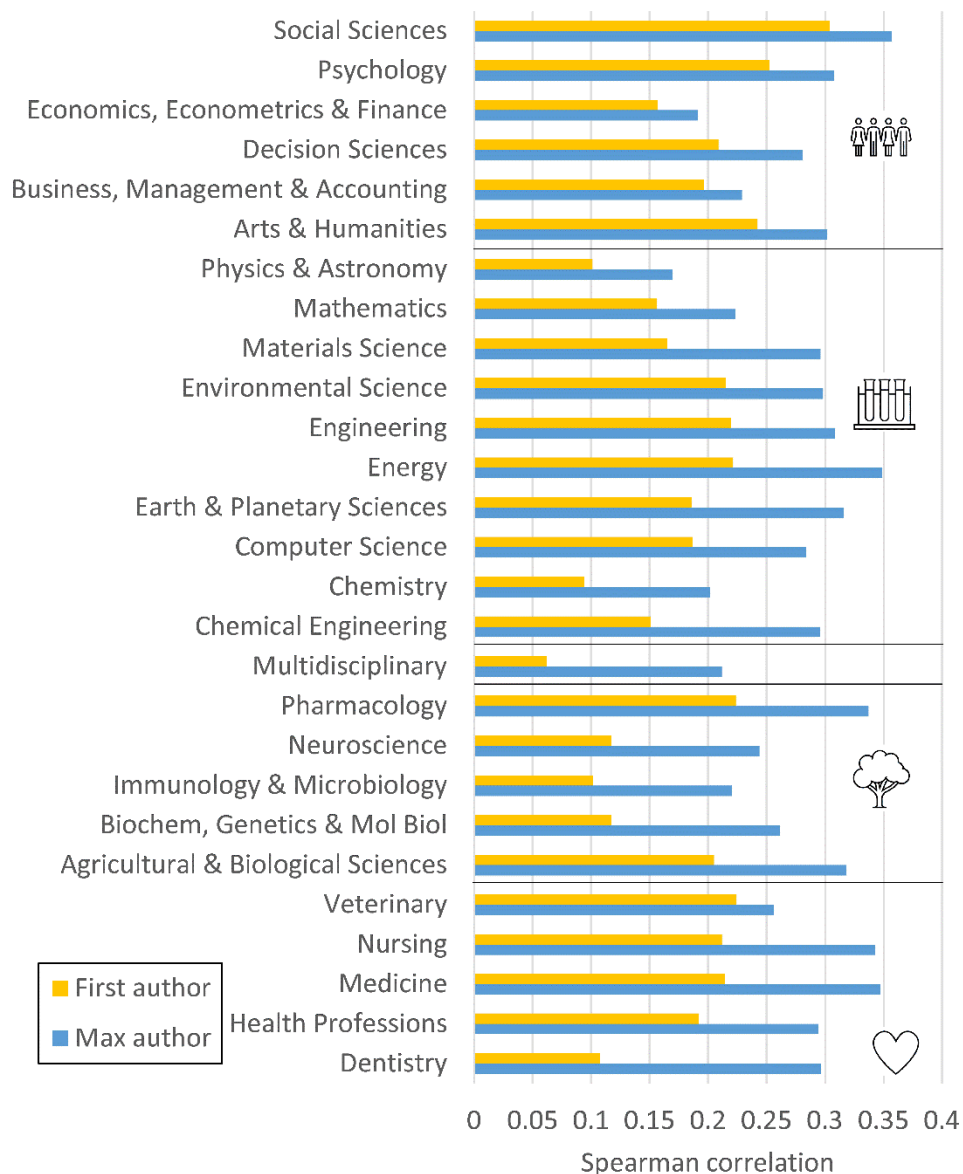


Figure 3. Average (over Scopus narrow fields) Spearman correlations between 2017 journal MNLCS and the number of articles 2014-2020 (full counting) by the **first author or authorship team maximum productivity (articles 2014-20)**.

### Associations between first author or team citation rate and journal impact

Correlations between first and maximum author average impact (MNLCS) and journal impact (MNLCS) were calculated in all narrow fields and averaged across broad fields to help identify trends in the results.

Both the first author average cited rate (MNLCS) and the team maximum cited rate have moderate or strong Spearman correlations with journal MNLCS in all broad fields (Figure 4). Of the two, the first author MNLCS tends to have a slightly higher correlation with journal MNLCS in all broad fields. The difference is highest in Veterinary. Thus, the results suggest that *higher cited authors or teams tend to publish in higher cited journals in all broad fields of science.*

There do not seem to be systematic differences between the four disciplinary clusters in the relationship between first author or team citation impact and journal citation impact, although the Health Sciences correlations tend to be lower than for the other three clusters.

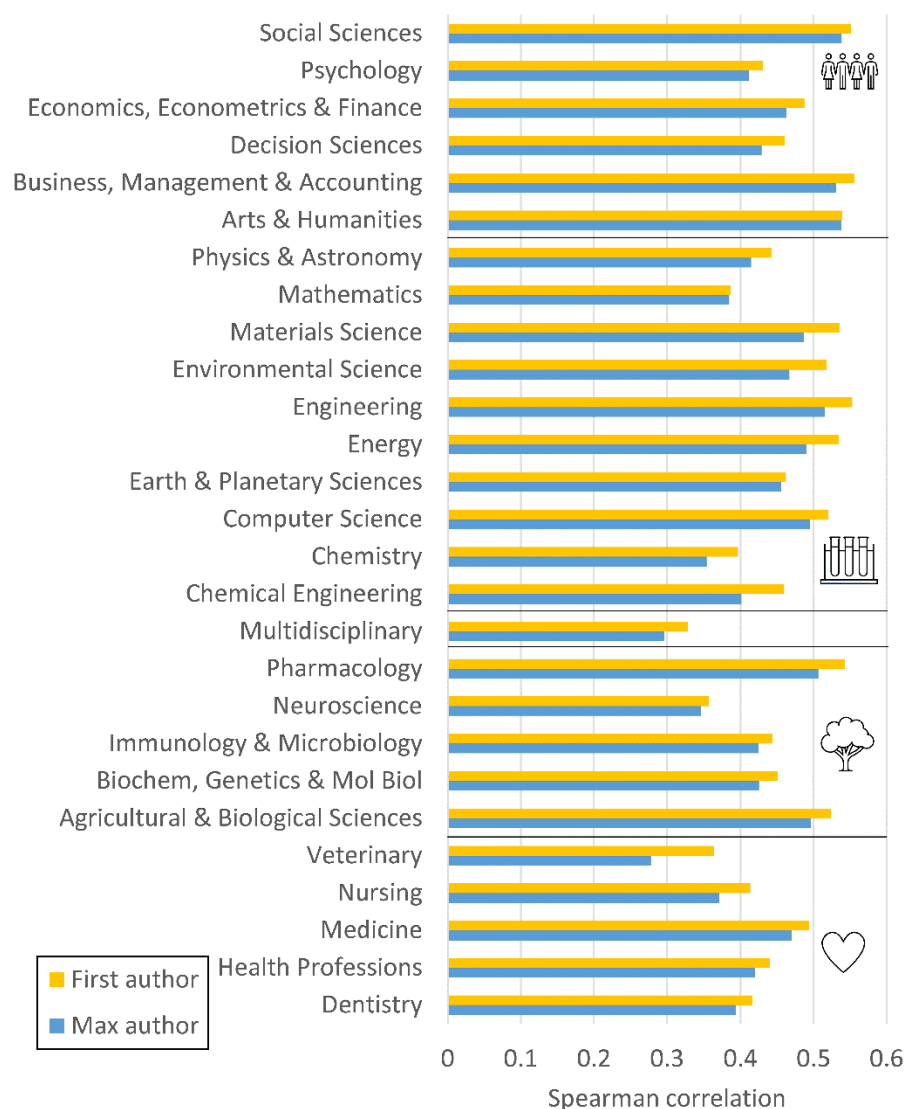


Figure 4. Average (over Scopus narrow fields) Spearman correlations between 2017 journal MNLCS and the 2014-2020 author MNLCS for the **first author or authorship team maximum average citation impact (2014-20)**.

### *First author and team productivity and impact considered simultaneously*

The correlation results above show associations but do not consider overlapping authorship properties (first author/team productivity/impact). This issue was addressed with regressions applied to single nations. Although restricted to domestic research from large English-speaking countries, the regression results allow the journal impact associations with first author and team productivity and impact to be simultaneously assessed in each broad field, also taking into account collaboration-related control variables. In theory, this could show if one of these were redundant (e.g., perhaps only the first author is relevant and the positive team correlations above are a side-effect of the first author). For the regressions, the benefits of international collaboration are irrelevant by design.

The regressions for domestic journal articles for large predominantly English-speaking countries only partly confirm overall the correlation results in this more restricted context, but with much larger disciplinary differences, including broad field exceptions. Results are shown here for the USA (Figure 5) and the UK (Figure 6), which have the most articles. Results for Australia, Ireland and New Zealand are available in the online supplement (<https://doi.org/10.6084/m9.figshare.19091687>).

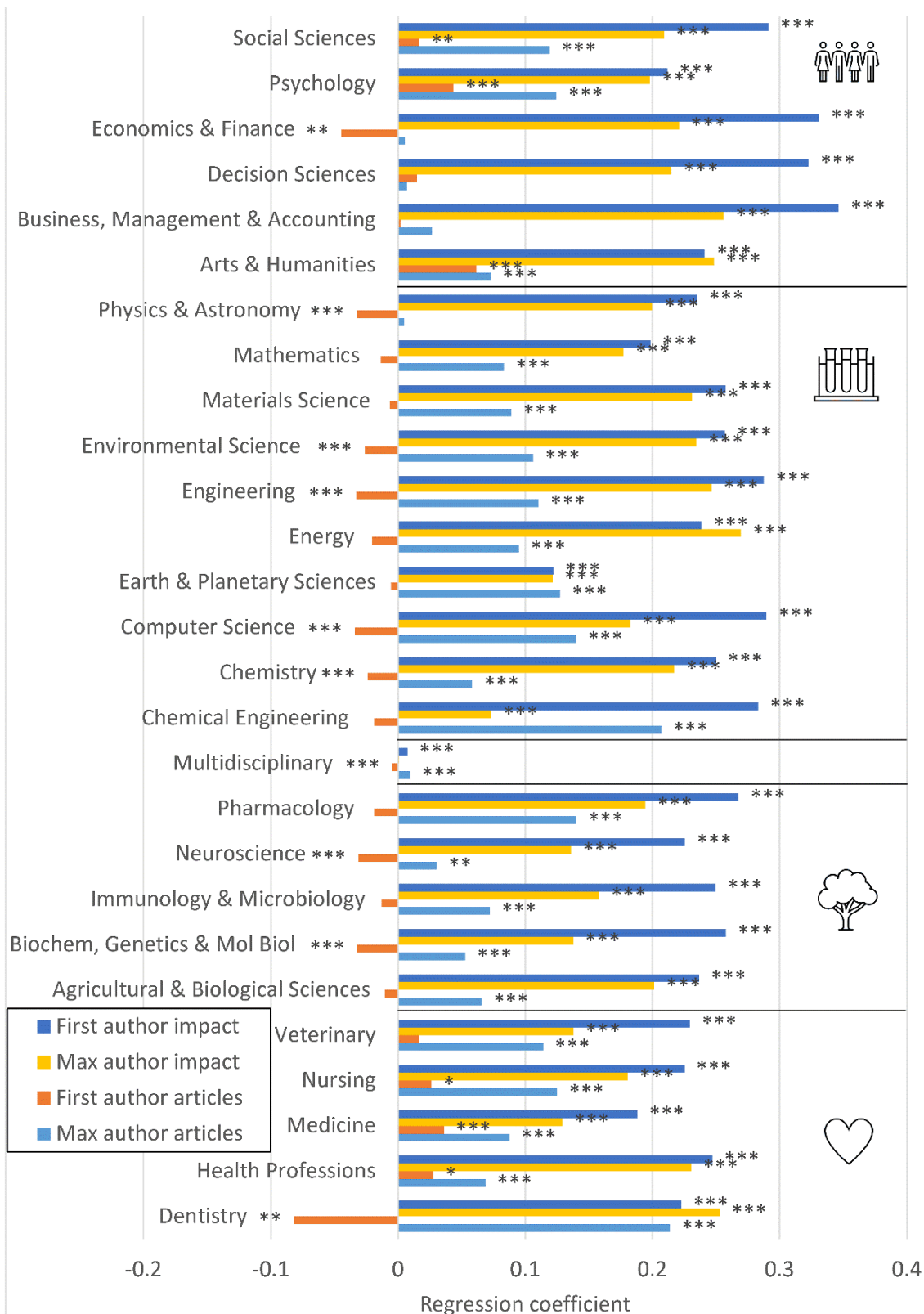


Figure 5. Standardised regression coefficients for the first author or team maximum productivity (number of Scopus-indexed journal articles 2014-20) and citation impact (MNLCS 2014-20) for **US** authors (regression dependant variable: 2017 journal MNLCS for domestic journal articles). There is one regression for each broad field and the four bars are for different coefficients from the same regression.

### First author and author team productivity vs. journal citation impact

For US domestic research, the first author productivity and team productivity components of the regression results (the lower two bars for each broad field in Figure 5) differ from the (international) correlations (Figure 3). Whilst a team maximum higher publishing volume tends to be an advantage (not always statistically significant), a higher publishing first author is – surprisingly - rarely an advantage, with the main exceptions being Psychology, Arts & Humanities (presumably due to the predominance of solo research) and Medicine (Figure 5). Surprisingly, a lower publishing first author (e.g., a PhD student) is an advantage in many fields, after factoring out team productivity and first author/team impact (Figure 5). A lower publishing first author is apparently particularly helpful for Dentistry, Economics and Finance, Physics and Astronomy, and Computer Science.

There are some broad disciplinary trends in these results, unlike most of the correlations. For the maximum author productivity regression coefficients, whilst all four disciplinary clusters have overlapping values, half of the Social Sciences & Humanities broad fields have lower values than the broad fields in the remaining three clusters, except the Physics and Astronomy broad field. This is not a strong pattern, however. For first author apparent contributions, there are clear disciplinary differences. The first author productivity contribution is negative throughout the physical sciences and life sciences but positive in all broad fields except one each in the Social Sciences and Humanities and the Health Sciences. Since a positive regression coefficient suggests that first author productivity is helpful for publishing in higher impact journals, this – again surprisingly - suggests that *less* productive first authors are more likely to get their work published in higher impact journals in the Physical Sciences and Life Sciences, at least in the USA. For example, this might mean that PhD work has a slightly better chance of being excellent than the papers of experienced researchers in these areas.

### First author and author team citation impact vs. journal citation impact

The author average citation impact (MNLCS) coefficients for the same set of regressions (the upper two bars for each broad field in Figure 5) are more uniform and fairly similar to the corresponding correlation results (Figure 4): a higher first author or team MNLCS is an advantage in all fields except Multidisciplinary for being published in a higher impact journal (Figure 5). The magnitude of the advantage is similar in many fields but there is not a consistent pattern about whether the first author or team maximum has the strongest association with the publishing journal MNLCS.

Comparing the standardised regression coefficients for author publishing productivity and author MNLCS, the MNLCS values tend to be higher and the combined effect of both first and maximum author MNLCS (since the effects are additive, unlike for correlation) shows that the collective MNLCS factor is much stronger than the collective author publishing productivity variables.

The four main disciplinary clusters mostly have values that overlap with each other, but there are some general trends, and particularly for first authors. The first author average impact contributions tend to be highest in the Social Sciences & Humanities and lowest in the Health Sciences. In contrast, the maximum author impact contributions tend to be weakest in the Life Sciences and Health Sciences. These patterns seem counterintuitive since journal impact is most relevant in medicine and the life and physical sciences, although it is also considered important in business and economics.

## The UK case

The results for the UK (Figure 6), broadly echo those of the US, but the patterns are less clear and less likely to be statistically significant. This is presumably due to less data. The main exception is that a lower publishing authorship team associates with higher journal impact in several fields, although the association is not statistically significant in any.

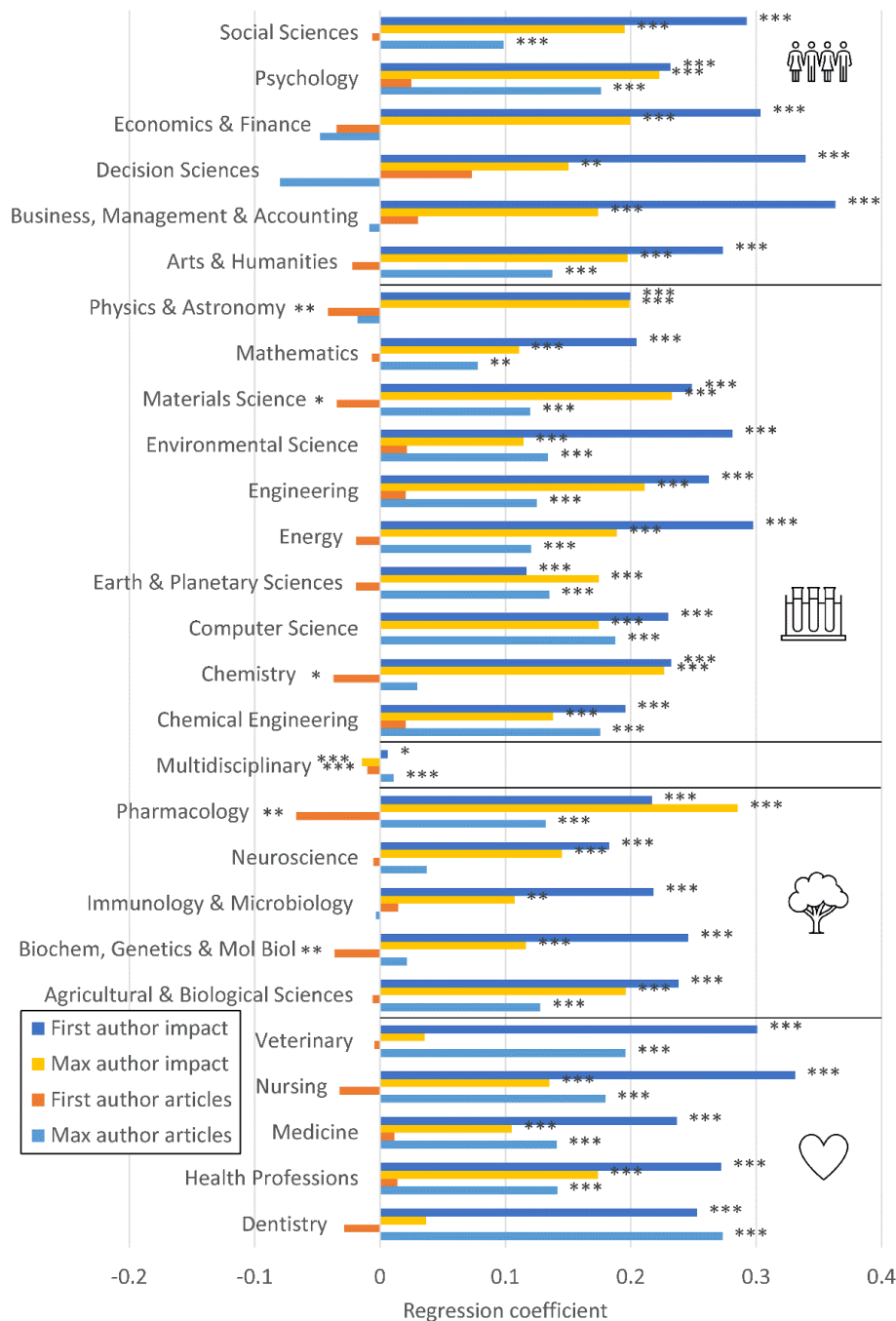


Figure 6. Standardised regression coefficients for the first author or team maximum productivity (number of Scopus-indexed journal articles 2014-20) and citation impact (MNLCS 2014-20) for UK authors (regression dependant variable: 2017 journal MNLCS for domestic journal articles). There is one regression for each broad field and the four bars are for different coefficients from the same regression.

## Discussion

The results are limited by the year chosen for analysis, the method of taking into account publishing careers (a 7-year window centred on the year analysed), the restriction to Scopus-indexed articles and Scopus-indexed broad and narrow field categories, the restriction to journal articles, and, more generally, the focus on publishing rather than other forms of academic work. The correlation results also do not consider the nature of the collaborations (e.g., single or multiple institutions), whereas the regression results only cover domestic research from large English-speaking nations (and mainly the USA for the greatest statistical power); thus, neither are fully satisfactory and the combination leaves gaps. Finally, none of the results imply causation. For instance, strong correlations between team size and journal citation impact might be partly due to funders financing larger teams rather than the intrinsic benefits of collaboration (Álvarez-Bornstein & Bordons, 2021).

### *First author or team productivity*

The results suggest that recruiting an authorship team member with more publications is far more helpful than the first author having written more publications, although both are less helpful than first or other authors tending to produce more cited work. This is evident from the correlations in Figure 3 (all research) and very clear in Figure 5 (US domestic research). In fact, the regression results suggest that, after factoring out collaboration factors and first author citation impact, an inexperienced first author is a small advantage in most broad fields.

No previous study seems to have contrasted first author and team productivity. Nevertheless, the results indirectly align with the finding that more prolific researchers tend to generate more citations per paper (Larivière & Costas, 2016; Sandström & van den Besselaar, 2016) in the sense that this shows a relationship between productivity and citation impact.

It is reasonable to believe that experience within a research team is helpful to get a study published in a high impact journal. Experience can help ensure that research is appropriately conducted and suitably framed for publication in a journal. Selecting an appropriate journal is presumably also helped by experience, given that it can be a difficult task. Of course, authors may also get help from experienced colleagues that are not co-authors, such as a departmental head.

It seems counterintuitive to believe that first author experience, even after taking other collaboration factors into account, is usually a disadvantage in getting an article published in a higher impact journal, despite the clear regression results. This seems likely to be an indirect effect, although the causes can only be speculated about. For example, perhaps PhD students are often first authors in many fields, and they produce good studies when supported by excellent supervisors (Thelwall, 2020c) and have more time to devote to a study than a busy academic. Alternatively, the result might be an artefact of the restriction to domestic research in the regressions, with more successful experienced researchers tending to collaborate internationally more (Abramo et al., 2019), partly taking themselves out of the dataset.

### *First author or team average citation impact*

The correlation and regression results suggest that publishing in a more cited journal almost equally associates with first authors or the maximum team author (defined as above) tending to publish more cited articles. Nevertheless, and in contrast the publication productivity



situation, the first author has a slightly stronger correlation and a moderately stronger regression coefficient than the team maximum.

No prior study seems to have investigated the relationship between average citations for authors and the citedness or journal impact of a particular article, with most career-level analyses analysing instead total citations or the h-index. Nevertheless, there are clearly differences between scholars in the average citation impact of their articles (Larivière & Costas, 2016; Mazloumian, 2012; Sandström & van den Besselaar, 2016), even after taking into account field and year differences (as the current paper does). Such differences in the correlations could also be a side effect of recognised substantial international differences in citations per paper (Elsevier, 2017), since authors in low-citation countries would presumably tend to have career-long lower citation rates than authors in high citation countries. The benefits of collaborating with experienced authors have also been recognised (Bu et al., 2018b; Thelwall, 2020a).

There are many reasons why a high citation track record for a first or maximum author could help an authorship team to publish in a higher impact journal. First, even after taking field difference in citation rates into account, individual researchers within a field may specialise in high citation topics and therefore tend to publish in higher citation journals associated with that topic. The extent to which this occurs would naturally vary between fields, being strongest in fields with many specialist journals and greatly differing citation rates between topics. Second, there may be relatively stable differences between scholars in their tendency to produce more impactful or otherwise higher quality (e.g., methodologically more robust, more innovative, or more influential) research, and there is some evidence for this for domestic research (Maflahi & Thelwall, 2021). Thus, a researcher that is skilled at producing impactful research would be expected, on average, to be more likely to get each article in a more cited journal.

The reason why the average citation rate of the first author is usually a better indicator than the team maximum average citation rate for the publishing journal citation impact is unclear. Speculatively, a senior researcher that tends to author highly cited articles may sometimes submit articles to low impact journals when they help an inexperienced first author (not necessarily a PhD student) or when they are the head of a laboratory that adds their name to all papers from their group, irrespective of contribution. On this basis, the first author may be more in control of the quality of the research that they submit, as well as where they submit it to.

### *Productivity vs. impact*

The results suggest that authorship experience (number of publications) is less important than average impact (citations per paper for the authors) for getting an article in a higher impact journal, irrespective of whether the first author or team is considered. This issue does not seem to have been investigated before. Even though productivity and citation impact are statistically related, it is not surprising that article citation impact associates more strongly with journal citation impact since both are citation averages.

### *Disciplinary differences*

Disciplinary differences are to be expected in any scientometric study because of the social and technical factors that influence scientific work vary substantially between fields (Whitley, 2000). It is nevertheless surprising that the differences within the four clusters examined (Social Sciences & Humanities; Physical Sciences; Life Sciences; Health Sciences) are in most

cases larger than differences between them. This suggests that relatively fine-grained details about how disciplines organise research influence the relationship between authorship and journal impact relatively strongly. Speculatively, such issues might include field norms about authorship order, the role of senior authors, expectations of PhD students and junior researchers, and the extent to which scholars contribute to different types of research or provide support to others' research projects.

The most substantial difference between the four disciplinary clusters is the negative contribution of first author productivity in the physical sciences and life sciences, suggesting that junior first authors or other rarely publishing authors are more likely to be published in higher impact journals than are senior or more frequent first authors, other factors (e.g., average citation impact) being equal in these two clusters. This may be a side effect of larger team sizes in the physical and life sciences. Although there are also large team sizes in the health sciences, there may be health-specific factors that undermine the large team effects, such as the need for more substantial ethical review and perhaps more careful publishing when human (or animal) health is involved.

Another cluster-level disciplinary difference is that first author average citation impact tends to be most important in the Social Sciences & Humanities and least in the Health Sciences. It is possible that the first author in the Social Sciences & Humanities has a greater role than in other areas and often tends to be the senior author because teams are quite small and therefore need less work to coordinate and fund. In contrast, the first author may be least important in health research because of the importance of complex funding, large teams, and ethical considerations.

Finally, the maximum citation rate in the team is most important in the Life and Health Sciences. In the Health Sciences this may be for the reasons given above, but it is not clear why the pattern would be different between the Life and Physical Sciences, both of which often rely on extensive funding and large teams needing coordination.

## Conclusions

The results show the following for domestic research in large English-speaking countries (with the strongest findings from the USA), in a way that is consistent with the correlations covering all countries:

- Articles with a first author or co-author tending to write higher cited research are moderately more likely to be published in a higher impact journal.
- Articles with an experienced first author or co-author (many publications) are slightly more likely to be published in a higher impact journal, but in most fields first author publishing productivity is a marginal disadvantage.
- Differences between the 26 Scopus Broad fields within of the four clusters (Social Sciences & Humanities; Physical Sciences; Life Sciences; Health Sciences) tend to be larger than differences between clusters, so fine-grained disciplinary differences are important for the relationship between authorship properties and journal impact.
- First author productivity associates with lower impact journals for the physical sciences and life sciences but higher impact journals in the Social Sciences and Humanities and the Health Sciences.
- In terms of journal impact, first author average citation rates tend to be most important in the Social Sciences & Humanities and least important in the Health Sciences.

- In terms of journal impact, team maximum average citation rates tend to be most important in the Social Sciences & Humanities and the Physical Sciences and least important in the Life and Health Sciences.

The results only partially cover scientific publishing because the correlation findings do not separate out the effects of the different variables and the regression results only cover domestic research from large predominantly English-speaking countries.

The correlation and regression results are consistent with, but do not prove, the idea that attracting highly cited team members to a research team would help to get the articles produced published in a higher impact journal. On average, higher impact journals are higher quality (Kelly et al., 2013; Mahmood, 2017; Steward & Lewis, 2010) and contain higher quality articles in some fields, so the same authorship factors may associate with the quality of an article. The associations have substantial disciplinary differences in strength, but they are present to some extent in all fields of science. This study was unable to account for all factors that may affect the results, so the evidence for the causation suggestion above is very weak.

The quantitative findings from this research need extensive qualitative follow-up to identify the reasons for the associations found. In the Discussion above, the speculative suggestions given are largely not evidence-based and smaller scale studies may help to give more concrete explanations.

If the causation explanation for the results is accepted as likely to be at least partly true, then the results underpin the importance of encouraging impactful researchers to help research teams to achieve the best results from their tasks. This seems to be more important than attracting experienced researchers to help with new teams.

## References

- Abramo, G., D'Angelo, C. A., & Di Costa, F. (2019). The collaboration behavior of top scientists. *Scientometrics*, 118(1), 215-232. <https://doi.org/10.1007/s11192-018-2970-9>
- Abramo, G., & D'Angelo, C. A. (2015). The relationship between the number of authors of a publication, its citations and the impact factor of the publishing journal: Evidence from Italy. *Journal of Informetrics*, 9(4), 746-761.
- Álvarez-Bornstein, B., & Bordons, M. (2021). Is funding related to higher research impact? Exploring its relationship and the mediating role of collaboration in several disciplines. *Journal of Informetrics*, 15(1), 101102.
- Amjad, T., & Munir, J. (2021). Investigating the impact of collaboration with authority authors: a case study of bibliographic data in field of philosophy. *Scientometrics*, 126(5), 4333-4353. <https://doi.org/10.1007/s11192-021-03930-1>
- Bu, Y., Ding, Y., Liang, X., & Murray, D. S. (2018a). Understanding persistent scientific collaboration. *Journal of the Association for Information Science and Technology*, 69(3), 438-448.
- Bu, Y., Murray, D. S., Xu, J., Ding, Y., Ai, P., Shen, J., & Yang, F. (2018b). Analyzing scientific collaboration with "giants" based on the milestones of career. *Proceedings of the Association for Information Science and Technology*, 55(1), 29-38.
- da Silva, J. A. T., Dobránszki, J., Van, P. T., & Payne, W. A. (2013). Corresponding authors: rules, responsibilities and risks. *Asian Australian Journal of Plant Science and Biotechnology*, 7(1), 16-20.

- Ding, A., & Herbert, R. (2022). Corresponding Authors: Past and Present How Has the Role of Corresponding Author Changed Since the Early 2000s? International Center for the Study of Research. <http://doi.org/10.2139/ssrn.4049439>
- Elsevier (2017). International comparative performance of the UK research base 2016. <https://www.elsevier.com/research-intelligence?a=507321>
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). The new production of knowledge: The dynamics of science and research in contemporary societies. Thousand Oaks, CA: Sage.
- Glanzel, W. (2002). Coauthorship Patterns and Trends in the Sciences (1980-1998): A Bibliometric Study with Implications for Database Indexing and Search Strategies. *Library Trends*, 50(3), 461-73.
- Grácio, M. C. C., de Oliveira, E. F. T., Chinchilla-Rodríguez, Z., & Moed, H. F. (2020). Does corresponding authorship influence scientific impact in collaboration: Brazilian institutions as a case of study. *Scientometrics*, 125(2), 1349-1369.
- Hall, K. L., Vogel, A. L., Huang, G. C., Serrano, K. J., Rice, E. L., Tsakraklides, S. P., & Fiore, S. M. (2018). The science of team science: A review of the empirical evidence and research gaps on collaboration in science. *American Psychologist*, 73(4), 532.
- Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26(1), 1-18.
- Katz, J., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics*, 40(3), 541-554.
- Kelly, A., Harvey, C., Morris, H., & Rowlinson, M. (2013). Accounting journals and the ABS Guide: a review of evidence and inference. *Management & Organizational History*, 8(4), 415-431.
- Kurmis, A. P. (2003). Understanding the limitations of the journal impact factor. *Journal of Bone and Joint Surgery*, 85(12), 2449-2454.
- Larivière, V., & Costas, R. (2016). How many is too many? On the relationship between research productivity and impact. *PLOS ONE*, 11(9), e0162709. <https://doi.org/10.1371/journal.pone.0162709>
- Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., & Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science*, 46(3), 417-435.
- Larivière, V., Gingras, Y., Sugimoto, C. R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66(7), 1323-1332.
- Larivière, V. (2012). On the shoulders of students? The contribution of PhD students to the advancement of knowledge. *Scientometrics*, 90(2), 463-481.
- Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation*, 11(1), 3-15.
- Levitt, J. M., & Thelwall, M. (2013). Alphabetization and the skewing of first authorship towards last names early in the alphabet. *Journal of Informetrics*, 7(3), 575-582.
- Levitt, J. M., & Thelwall, M. (2016). Long term productivity and collaboration in information science. *Scientometrics*, 108(3), 1103-1117.
- Maflahi, N. & Thelwall, M. (2021). Domestic researchers with longer careers generate higher average citation impact but it does not increase over time. *Quantitative Science Studies*, 2(2), 560–587. [https://doi.org/10.1162/qss\\_a\\_00132](https://doi.org/10.1162/qss_a_00132)

- Mahmood, K. (2017). Correlation between perception-based journal rankings and the journal impact factor (JIF): a systematic review and meta-analysis. *Serials Review*, 43(2), 120-129.
- Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & López-Cózar, E. D. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics*, 126(1), 871-906.
- Mattsson, P., Sundberg, C. J., & Laget, P. (2011). Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position. *Scientometrics*, 87(1), 99-105.
- Mazloumian, A. (2012). Predicting scholars' scientific impact. *PloS One*, 7(11), e49246.
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213-228.
- Mongeon, P., Smith, E., Joyal, B., & Larivière, V. (2017). The rise of the middle author: Investigating collaboration and division of labor in biomedical research using partial alphabetical authorship. *PloS One*, 12(9), e0184601.
- Nicolaisen, J., & Frandsen, T. F. (2015). The focus factor: a dynamic measure of journal specialisation. *Information Research: An International Electronic Journal*, 20(4), n4.
- Pölonen, J., Guns, R., Kulczycki, E., Sivertsen, G., & Engels, T. C. (2021). National Lists of scholarly publication channels: an overview and recommendations for their construction and maintenance. *Journal of Data and Information Science*, 6(1), 50-86. <https://doi.org/10.2478/jdis-2021-0004>
- Rousseuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., & Maechler, M. (2021). robustbase: Basic Robust Statistics. R package version 0.93-9. <https://cran.r-project.org/web/packages/robustbase/robustbase.pdf>
- Sandström, U., & van den Besselaar, P. (2016). Quantity and/or quality? The importance of publishing many papers. *PLoS One*, 11(11), e0166149. <https://doi.org/10.1371/journal.pone.0166149>
- Scopus (2020). What is the complete list of Scopus Subject Areas and All Science Journal Classification Codes (ASJC)? [https://service.elsevier.com/app/answers/detail/a\\_id/15181/supporthub/scopus/~what-is-the-complete-list-of-scopus-subject-areas-and-all-science-journal/](https://service.elsevier.com/app/answers/detail/a_id/15181/supporthub/scopus/~what-is-the-complete-list-of-scopus-subject-areas-and-all-science-journal/)
- Seglen, P. O. (1998). Citation rates and journal impact factors are not suitable for evaluation of research. *Acta Orthopaedica Scandinavica*, 69(3), 224-229.
- Steward, M. D., & Lewis, B. R. (2010). A comprehensive analysis of marketing journal rankings. *Journal of Marketing Education*, 32(1), 75-92. <https://doi.org/10.1177/0273475309344804>
- Thelwall, M. (2017). Three practical field normalised alternative indicator formulae for research evaluation. *Journal of Informetrics*, 11(1), 128-151. <https://doi.org/10.1016/j.joi.2016.12.002>
- Thelwall, M. (2020a). Author gender differences in psychology citation impact 1996-2018. *International Journal of Psychology*, 55(4), 684-694.
- Thelwall, M. (2020b). Large publishing consortia produce higher citation impact research but co-author contributions are hard to evaluate. *Quantitative Science Studies*, 1(1), 290-302.

- Thelwall, M. (2020c). Author gender differences in psychology citation impact 1996-2018. *International Journal of Psychology*, 55(4), 684–694.  
<https://doi.org/10.1002/ijop.12633>
- Thelwall, M. & Maflahi, N. (2020). Academic collaboration rates and citation associations vary substantially between countries and fields. *Journal of the Association for Information Science and Technology*, 71(8), 968–978.  
<https://doi.org/10.1002/asi.24315>
- Uwizeye, D., Karimi, F., Otukpa, E., Ngware, M. W., Wao, H., Igumbor, J. O., & Fonn, S. (2020). Increasing collaborative research output between early-career health researchers in Africa: lessons from the CARTA fellowship program. *Global Health Action*, 13(1), 1768795.
- van Rijnsoever, F. J., & Hessels, L. K. (2021). How academic researchers select collaborative research projects: a choice experiment. *The Journal of Technology Transfer*, 46(6), 1917-1948.
- Wakeling, S., Spezi, V., Fry, J., Creaser, C., Pinfield, S., & Willett, P. (2019). Academic communities: The role of journals and open-access mega-journals in scholarly communication. *Journal of Documentation*, 75(1), 120-139.  
<https://doi.org/10.1108/JD-05-2018-0067>
- Wagner, C. S., Whetsell, T. A., & Mukherjee, S. (2019). International research collaboration: Novelty, conventionality, and atypicality in knowledge recombination. *Research Policy*, 48(5), 1260-1270.
- Waltman, L., & Traag, V. A. (2020). Use of the journal impact factor for assessing individual articles need not be statistically wrong. *F1000Research*, 9.  
<https://f1000research.com/articles/9-366>
- Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, 94(3), 851-872.
- Whitley, R. (2000). *The intellectual and social organization of the sciences*. Oxford, UK: Oxford University Press.
- Yu, J., & Yin, C. (2021). The relationship between the corresponding author and its byline position: An investigation based on the academic big data. *Journal of Physics*, 1883(1), paper 012129.