
Article:
Thelwall, M. orcid.org/0000-0001-6065-205X, Kousha, K., Abdoli, M. et al. (4 more authors) (2023) Why are coauthored academic articles more cited: higher quality or larger audience? Journal of the Association for Information Science and Technology, 74 (7). pp. 791-810. ISSN 2330-1635

https://doi.org/10.1002/asi.24755

Received: 11 December 2022 | Revised: 1 March 2023 | Accepted: 11 March 2023

DOI: 10.1002/asi.24755

RESEARCH ARTICLE

JASIST WILEY

Why are coauthored academic articles more cited: Higher quality or larger audience?

Mike Thelwall | **Kayvan Kousha** | **Mahshid Abdoli** | **Emma Stuart** | **Meiko Makita** | **Paul Wilson** | **Jonathan Levitt**

Statistical Cybermetrics and Research Evaluation Group, University of Wolverhampton, Wolverhampton, UK

Correspondence
Mike Thelwall, Statistical Cybermetrics and Research Evaluation Group, University of Wolverhampton, Wolverhampton, UK.
Email: m.thelwall@wlv.ac.uk

Funding information
Research England, Scottish Funding Council, Higher Education Funding Council for Wales, and Department for the Economy, Northern Ireland

Abstract

Collaboration is encouraged because it is believed to improve academic research, supported by indirect evidence in the form of more coauthored articles being more cited. Nevertheless, this might not reflect quality but increased self-citations or the "audience effect": citations from increased awareness through multiple author networks. We address this with the first science wide investigation into whether author numbers associate with journal article quality, using expert peer quality judgments for 122,331 articles from the 2014–20 UK national assessment. Spearman correlations between author numbers and quality scores show moderately strong positive associations (0.2–0.4) in the health, life, and physical sciences, but weak or no positive associations in engineering and social sciences, with weak negative/positive or no associations in various arts and humanities, and a possible negative association for decision sciences. This gives the first systematic evidence that greater numbers of authors associates with higher quality journal articles in the majority of academia outside the arts and humanities, at least for the UK. Positive associations between team size and citation counts in areas with little association between team size and quality also show that audience effects or other nonquality factors account for the higher citation rates of coauthored articles in some fields.

1 | INTRODUCTION

A key factor behind much science policy is encouragement for increased collaboration (Lee & Bozeman, 2005; Ubfal & Maffioli, 2011). For example, European Union research funding is often predicated on the participation of at least three different countries, Marie Curie and Fulbright scholarships fund mobility, and large-scale instrumentation requires extensive funding and cooperation (D'ippolito & Rüling, 2019). This policy is underpinned by historically influential theory arguing that big science or interdisciplinary collaboration are necessary to address complex societal challenges (e.g., de Solla Price, 1963; Gibbons et al., 1994; Ziman, 2000), and the importance of academic teamwork is recognized by a dedicated research field about team science (Hall et al., 2018).

The value of collaboration is supported by a substantial body of empirical research showing that articles tend to be more cited when they have more authors (e.g., most studies listed in Shen et al., 2021), although

the science-wide citation advantage of coauthorship decreased 1900–2011 (Larivière et al., 2015). Nevertheless, it has been argued that coauthored articles tend to be more cited mainly because the additional authors attract more interest through their personal networks, particularly when multiple countries are represented (Deichmann et al., 2020; Ding et al., 2021; Wagner et al., 2019), rather than coauthored studies tending to be of better quality. There is also evidence that more coauthored articles attract more self-citations (Glänzel & Thijs, 2004; Lin & Huang, 2012; Shen et al., 2021), another nonquality link between collaboration and citation counts. Moreover, solo authorship is valued in the arts and humanities (e.g., Hansson et al., 2021) so it is unclear whether coauthorship would have a similar association with quality in all fields. As reviewed below, two previous partial empirical analyses of the relationship between coauthorship and research quality (Bornmann, 2017; Franceschet & Costantini, 2010) have found weak relationships but robust science wide studies are needed to give detailed information about disciplinary differences. For research evaluation considerations, it is also important to assess whether more coauthored articles tend to be more cited in any fields for reasons that cannot be explained by the quality of the research produced, potentially due to nonresearch factors such as the "audience effect" and self-citations. This knowledge will help decision makers evaluating the effectiveness of team science with the aid of citation analysis.

This article reports the first science wide empirical analysis of the relationship between research quality and the number of coauthors of journal articles, filling the gaps identified above. It follows a previous similar study of mixed output types (Franceschet & Costantini, 2010) with methodological improvements that undermined the prior study (see below), a science wide scope, eight times more outputs, 15 years more recent publications, finer grained and multiple field classifications (34, 27, and 22 fields rather than 14), and evidence of statistical associations that is comparable between fields (correlations rather than group-based chi square tests). It addresses the following research questions.

- *RQ1: In which fields does coauthorship associate with higher research quality for journal articles?* This is the major policy-relevant issue because negative answers for any fields remove one strong part of the case for funders and research managers to encourage collaboration in them.
- *RQ2: What is the overall relationship (i.e., graph shape) between the number of journal article authors and research quality?* This adds depth to RQ1 by giving more detailed information than a binary yes/no. For example, the overall relationship might be U-shaped (solo authors and large teams write the best articles), so the overall correlation would be misleading.
- *RQ3: In which fields does journal article coauthorship associate with more citations but not higher research quality?* This would give the strongest evidence yet that the hypothesized audience effect (including other nonquality factors associating with citations, such as self-citations) exists, which would imply that research with more coauthors has an unfair advantage with citation indicators in those fields.

2 | BACKGROUND

Scientific collaboration has increased over time, encouraged by systemic factors, such as economic policies and incentives. The choice of collaborators is influenced by geopolitics, historical factors, shared languages (Luukkonen et al., 1992). More specifically, coauthorship tends to be used to acknowledge involvement in a study, although the nature and extent of contributions varies between fields (Larivière et al., 2016), and there are exceptions like ghost and gift authorship (Gülen et al., 2020). Collaboration does not necessarily lead to coauthorship of a journal article, but the two are conflated here for simplicity. The reasons why researchers may decide to collaborate include social pleasure, PhD supervision, junior research visits, funding, and access to equipment (Katz & Martin, 1997; Melin, 2000; Shen et al., 2017; Sonnenwald, 2007; Ubfal & Maffioli, 2011). Researchers may also collaborate to learn from each other or to provide complementary skills (e.g., statistics and survey design). Coauthorship may be effectively mandatory, for research funding or PhD supervision, or optional, as when two mathematicians take turns to tackle a problem of shared interest. Thus, coauthorship covers a range of different phenomena. The question of whether collaboration is advantageous overall is therefore asking whether the contexts in which it is either advantageous or necessary for high-quality research are more prevalent than the contexts in which it is disadvantageous or necessary for low-quality research. On this basis, there can be no overall theory of collaboration advantages.

2.1 | Coauthorship and citation counts

Positive associations between author numbers and citation-based indicators have been found in many contexts, always grouping all types of coauthorship together, and this section briefly reviews relevant large-scale

studies. In general, coauthorship has been found to associate with higher citation counts in most contexts analyzed, but with exceptions covering whole countries, whole disciplines, and country/discipline combinations. These could be due to differing combinations of types of collaboration. For example, the results might be weaker in countries/fields/eras when collaboration was not incentivized with funding, was primarily domestic, or was primarily for PhD research.

For Web of Science articles 2000–2009 in Biology and Biochemistry, Chemistry, and Social Sciences, there is a positive correlation between author numbers and citation counts (Didegah & Thelwall, 2013). For Italian-authored research 2004–2010, linear regressions revealed a positive association between the number of authors and article citations (percentiles within Italy) for all areas examined: Biology; Biomedical research; Chemistry; Clinical medicine; Earth and space sciences; Economics; Engineering; Law, political and social sciences; Mathematics; Multidisciplinary sciences; Physics; Psychology (Abramo & D'Angelo, 2015). The association was strongest in Clinical Medicine and weakest in Engineering. For South African papers from 2000, 2003, and 2005, collaborative articles were more cited than solo articles in most areas of academia, but the opposite was true in Psychiatry, Biochemistry, Agriculture, and Material Science (Sooryamoorthy, 2009). The latter fields are not exceptions in other studies. For the nine countries with the most articles in Scopus, domestic collaborative research (all authors from the same country) was more cited than solo research except in Russia (Thelwall & Sud, 2016). The same study found that collaboration was associated with more citations in arts and humanities much more strongly than in business, chemistry, and pharmaceutics. A larger-scale follow-up study of 10 countries (not including Russia) found the weakest association between citations and author numbers to be in China (Thelwall & Maflahi, 2020). It also found a general trend for domestic research to be more cited when it had multiple authors, although there were exceptions, such as business research in Germany. International examples of situations without an association between citations and collaboration include finance 1987–1991 (Avkiran, 1997). Thus, while greater numbers of authors associates with higher citation rates in most countries and fields, there are exceptions for unknown reasons.

There are multiple possible causes of collaborative research tending to be more cited that are unrelated to quality. One simple factor for which there is some empirical evidence is self-citation: there is a trend for articles with more authors to attract a higher proportion of self-citations (Glänzel & Thijs, 2004; Lin & Huang, 2012; Shen et al., 2021). This increases the possibility that the coauthored articles might not be higher quality despite being more cited. Related to this, the "audience effect" is that each author may bring an "audience" of people that know them or that are interested in their work, so larger teams may generate more readers and hence more citations from a fraction of these. This has also been called an "advertising factor" (Franceschet & Costantini, 2010). This seems especially likely for international coauthorship reaching different countries (Lancho Barrantes et al., 2012; Schmoch & Schubert, 2008; Wagner et al., 2019) since articles attract disproportionately much interest from their authors' nations (Thelwall & Maflahi, 2015). Finally, at a statistical level, the presence of a few articles from an influential large, highly funded team (Thelwall, 2020) can induce a positive correlation in a set of articles that otherwise would not show a trend.

2.2 | Coauthorship and research quality

Research quality is judged during peer review before publication, national research evaluations, post publication informal evaluations by readers, and job applications. Peer review has become the standard method of assessing the quality of journal articles, particularly prior to publication (Benos et al., 2007). There are three generally agreed main components of research quality: rigor, novelty/originality, and significance to science or society (Langfeldt et al., 2020). These dimensions vary in how they are assessed between fields (e.g., REF, 2020), despite being relevant to all.

There are many reasons why articles with a greater number of authors tend to be better in some aspect of quality, at least in some disciplines, and some reasons that point in the opposite direction (Beaver, 2013; Hall et al., 2018). The greater expertise brought by a team of people might allow their work to be higher quality overall—for example because each is a specialist for their task—or they may be able to tackle more important problems with larger or more interdisciplinary teams and more resources (Gibbons et al., 1994). Conversely, collaboration could reduce the strength of the weakest link in a team or the chance of radical discoveries by maverick scientists (Hull, 2010). There is no single cause of the collaboration citation advantage because of the variety of reasons for collaboration, disciplinary differences in research types, varied incentives to collaborate, and multiple benefits from collaboration (Van Rijnsoever & Hessels, 2011).

In some biological (Cantor et al., 2010) and medical research (e.g., vaccine trials), increasing the sample size for human subjects research is expensive and complex because it requires the work of more people to be

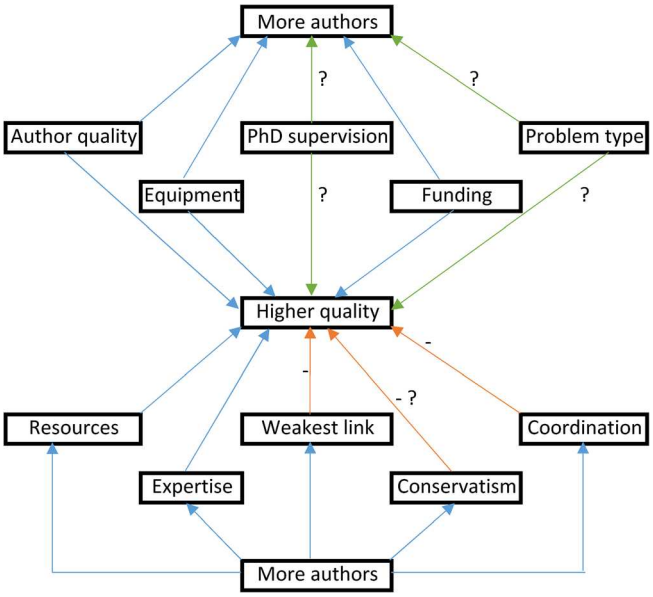

FIGURE 1 Some factors that might directly or indirectly contribute to the relationship between collaboration and research quality in a field. Blue arrows indicate a positive relationship, orange negative, and green unknown or variable between fields

coordinated. Thus, larger-scale studies are both harder to achieve and more likely to generate positive results. Other factors being equal, increased research quality (particularly in terms of significance), can be expected with larger numbers of coauthors in this case. In contrast, some arts and humanities work might primarily require the deep and creative thought of an individual scholar, who may be skeptical about the value of collaboration (e.g., as mentioned in Real, 2012; the importance of [usually solo] monographs in Shaw et al., 2022). Similarly, in some cases a collaboration might indicate that a senior scholar had delegated some of the work, such as the literature review or one experiment, to a junior colleague, perhaps affecting its quality.

From a different perspective, some research goals, such as vaccine development (Gilbert & Green, 2021) and high energy physics experiments (Perkins, 2000), are fundamentally big science endeavors that require large-scale coordination of different specialisms. These are impossible to conduct on a small scale, so if their outputs are judged to be high quality—for example as relatively unique contributions to science—then this alone would generate a statistical association between team size and research quality.

Funding is important in many research fields, and a better funded study might operate on a larger scale, producing higher quality, more useful results. If part of the funding is spent on creating a larger team, then this would create an association between team size and quality. Funding in medicine associates with higher rates of citation, partly because of the larger team sizes of funded research (Yan et al., 2018).

From yet another perspective, interdisciplinary research must include a relatively unusual combination of input methods or expertise (e.g., Wagner et al., 2019) and therefore seems likely to be judged to have higher methodological originality and may also be more likely to generate societal impacts (e.g., Gibbons et al., 1994; Ignaciuk et al., 2012).

Figure 1 illustrates the complexity of the relationship between collaboration and research quality, summarizing some of the above points. While more authors would bring more expertise and resources, which should increase research quality, they may also weaken the weakest link, increase conservatism, and bring coordination problems, which may detract from research quality. On the other hand, expensive equipment and large funding pots might generate large teams and good research (because better resourced). Also, a better researcher might produce better research and attract more coauthors. For PhD supervision, the research might tend to be higher or lower quality in some fields, and the PhD student may tend to work in fields that are larger or smaller than the field average, so this relationship may well vary between fields. The current paper only investigates the overall relationship between author numbers and publication quality, but this diagram emphasizes the causal factors that may underly this relationship.

2.3 | Empirical evidence of relationships between coauthorship and research quality

Given the complex multiple factors at work, empirical evidence is needed to identify the fields in which there is a clear association between collaboration and research quality. There is limited indirect evidence of this from a few small-scale studies. For example, an old study found that articles with more authors were less likely to be rejected during peer review in a psychology journal (Presser, 1980). From a different perspective, graduate library and information science students tended to find collaborative journal articles more useful for their course than solo authored journal articles (Finlay et al., 2012), addressing the significance dimension of quality. Finally, if journal impact is an acceptable proxy for research quality, then New Zealand biomedical scientists tend to produce higher quality research when they collaborate (He et al., 2009).

Two studies have investigated the relationship between author numbers and peer reviewed journal article quality on a larger scale, neither of which were

science wide. An analysis of 16,554 biology and medicine articles from before 2014 found that the (field and year normalized) citation advantage of collaboration could not be explained by the quality of collaborative articles, as judged by F1000Prime (now Faculty Opinions) reviewers. There was a very weak Spearman correlation (0.09) between article quality and the number of authors per paper (Bornmann, 2017). The weak relationship may have been due to the relatively ad hoc nature of the reviewing, at least compared to the study below.

A large-scale multidisciplinary study of 15,301 Italian research outputs 2001–2003 (10,665 journal articles in the Web of Science used for the citation analysis component, including few in the arts and humanities: 2 Law articles, 5 for Philological-literary Sciences, Antiquities and Arts, and 23 for Political and Social Sciences) rated by expert peer review analyzed 14 out of 20 categories, ignoring six interdisciplinary fields. It found that outputs of all types (combined rather than analyzed separately) with more authors tended to have higher quality ratings in most fields: Physics; Earth Sciences; Biology, Medical Sciences; Agricultural Sciences and Veterinary Medicine; Philological-literary Sciences, Antiquities and Arts; History, Philosophy, Psychology, and Pedagogy; Economics and Statistics; and Political and Social Sciences. There was not a clear positive relationship in a few fields: Chemistry; Mathematics and Computer Sciences; Civil Engineering and Architecture; Industrial and Information Engineering; and Law (Franceschet & Costantini, 2010). Unfortunately, one of the four quality components that reviewers were asked to assess was "internationalization," which conflicts with an attempt to use the results to investigate the relationship between quality and coauthorship since reviewers may have regarded international coauthors as evidence of internationalization, giving higher scores for this. Similar authorship team sizes were grouped together for this analysis (e.g., 1, 2, >2 for Mathematics and Computer Sciences), which may have hidden finer-grained relationships, and only six of the differences were statistically significant. Moreover, combining substantially different output types (e.g., books and journal articles), which probably have different average quality scores and coauthorship team sizes, undermines the validity of the results. For example, in Law, books may have fewer authors and higher quality scores than journal articles because monographs are core to the humanities. Thus, this study, despite being the largest so far, does not give strong evidence the relationship between research quality and team size for journal articles (because this was not checked directly) or for all output types (because of output mixing effects). Larger-scale studies with more robust methods and science wide data are needed to give plausible evidence. More fine-grained classifications are also needed to reduce the risk that correlations are due to disciplinary differences in the results (e.g., higher collaboration fields tending to produce higher quality scores than lower collaboration fields that they are mixed with [or vice versa], as could be the case for the Mathematics and Computer Science category).

3 | METHODS

3.1 | Data

We obtained provisional quality scores for 148,977 journal articles submitted to the UK REF2021 from the REF team under a confidentiality agreement as part of a research contract in March 2022, excluding submissions from the University of Wolverhampton for privacy reasons. REF journal articles must be primary research rather than reviews and are the self-selected best 1–5 outputs of all active UK researchers first published between 2014 and 2020. We were not given information about the other output types submitted by researchers (e.g., books, software, performances). The data includes scores on a 4-point scale for overall quality: 1*, 2*, 3*, or 4*. Articles were scored 0 if they did not qualify for review or were judged not to be research. The scores are allocated by 34 subpanels of 1,000+ mostly senior subject experts (listed here: REF2021, 2021b), usually with at least two per output. Subpanels met for quality control purposes and evaluators were given 17 pages of guidelines about how to evaluate the quality of the research overall and in their area (REF, 2020, pp. 34–51). In addition, there was preliminary training, and substantial norm-referencing within and between UoAs. Based on REF2014 data and inflation, the output scoring in REF2021 may have cost £20.6 million (12% Bank of England cumulative inflation added to £18.4 million for REF2014 based on 40% (Figure 5) of the REF2014 output assessment £46 million cost [p. 8]: Technopolis, 2015), underlining the care given to the quality scoring. Nevertheless, the scores are imperfect because the panel members do not cover all academic specialties.

We discarded all 318 articles with score 0 since at least some were high-quality articles that had been judged out of scope for authorship reasons or type reasons (judged to be a review). The remaining articles included many duplicates: articles submitted by multiple authors. We eliminated these duplicates separately within each analysis unit (UoA or main panel), retaining the median score when an article had been given different scores within the unit analyzed (chosen at random when there were two medians).

We matched the REF articles with Scopus articles published 2014–20 primarily by DOI but also by title and authors for additional matching (similar to Bornmann, 2017), with human checks of these additional results, giving 997 new matches. We discarded non-matching REF articles. We used the Scopus matches to record the number of authors for each article, as listed in Scopus, and for the Scopus broad fields of each article. Scopus assigns articles to broad fields usually based on the publishing journal. When a journal was assigned to multiple fields then all its articles were also assigned to all those fields. This produced 122,331 REF papers scoring 1* to 4* and matching a Scopus record 2014–20.

We also searched for the 122,331 articles that matched with Scopus in the scholarly search engine Dimensions.ai (Hook et al., 2018) through DOIs, although it was not practical to also check by title for the missing articles. All articles for Dimensions.ai also had to be in Scopus to use the Scopus author counts, which seemed to be the most reliable. This restriction is not a major problem because Dimensions indexes 84% of Scopus (Martín-Martín et al., 2021). We used the Dimensions records only for its top-level Field Of Research (FOR) codes, which are assigned to articles on an individual basis (Dimensions.ai, 2021), with each article typically assigned multiple broad and/or narrow codes. We excluded articles that had not been assigned a FOR code (because the Dimensions field classification AI software had not reached its probability threshold).

After these processing steps, there were 134,801 articles in Dimensions broad fields, multiply counting articles with more than one broad FOR code, 122,331 articles in UoAs, multiply counting articles submitted by authors to more than one UoAs, and 201,635 articles in Scopus, multiply counting articles in more than one Scopus broad field.

For the second research question, we obtained raw Scopus citation counts for all the articles analyzed from January 2021 but these would not be appropriate for comparisons because (a) the data is from multiple years and (b) UoAs combine multiple fields. Thus, we transformed the citation counts into field normalized scores, the Normalized Log-transformed Citation Score (NLCS) (Thelwall, 2017). For this, we first log transformed all citation counts with $\ln(1 - x)$ to reduce skewing and support the calculation of more precise averages. Next, we calculated the average log-transformed citation count for every Scopus narrow field and year (326 fields $\times$ 7 years = 2,282 averages, with some empty field/year combinations), based on all standard journal articles, not just the REF2021 articles. We then calculated the NLCS for each article as its log-transformed citation count $\ln(1 + x)$ divided by the average for its field and year. Articles in multiple fields were divided by the average of the averages for each of the fields instead. The NLCS for each article is 1 if the article had the average number of citations for all articles in its field(s). Scores greater than one indicate above average citation impact and scores below one indicate below average citation impact, irrespective of field(s) and year. By design, is fair to compare NLCSs for articles in different fields and years since each NLCS calculation is norm referenced only against its field(s) and year.

We included the Scopus citation counts for all years (rather than excluding the most recent years to give a citation window of 2–3 years). This was done so that the dataset for citations matched the dataset for quality in terms of articles and for confidence interval width calculations. This has the undesirable side-effect that some articles would have a citation window as short as a month, although the average would be 3.5 years. This adds noise to the data rather than systematic bias, however, and the amount of noise is reduced by the use of the log-based NLCS citation indicator. Moreover, the effective citation window will have been increased by early online publication for many, since the REF mandates open access publication as soon as possible and within three months of initial acceptance, unless an exception applies.

3.2 | Analysis

We assessed the relationship between the number of authors and the quality of an article (1* to 4*) using Spearman correlations. This approach was used instead of regression to investigate the overall pattern rather than the contribution of coauthorship relative to other factors or the effect of different types of collaboration. We log transformed author numbers with $\ln(1 + x)$ and Pearson correlations were also used with the log-transformed data. Pearson correlations are more problematic because the 4-point REF scoring system is an ordered set of qualitative categories rather than a scale. The results of the two types of correlation were similar so we only report the Spearman correlations for simplicity. 95% confidence intervals were calculated using the Fisher (1915) transformation.

We calculated the same correlations for author numbers and citation counts. The two sets of Spearman correlations are not directly comparable because of the large difference in granularity between NLCS and quality scores, which affects the magnitude of a correlation for a similar underlying relationship strength (Thelwall, 2016).

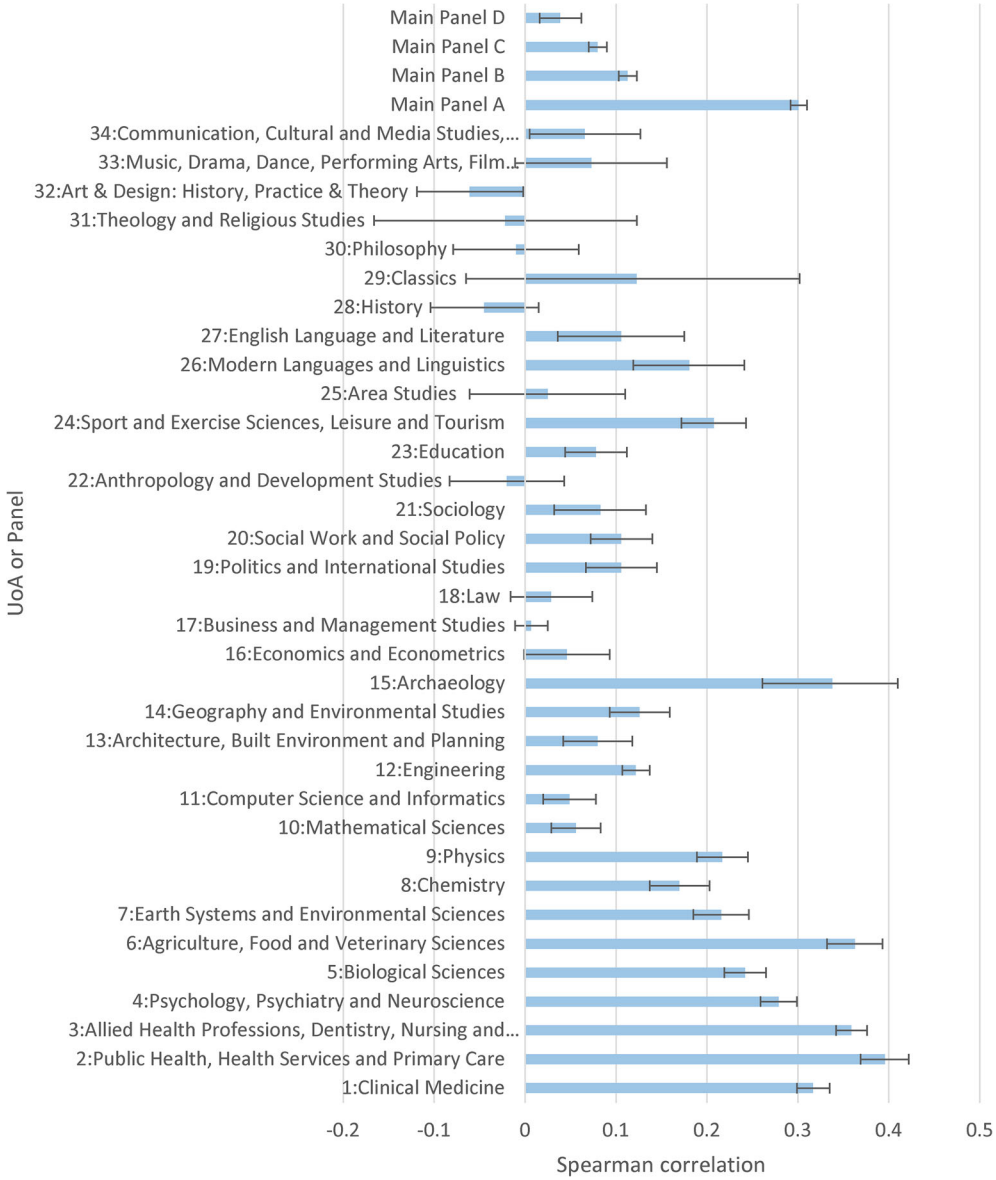

FIGURE 2 Spearman correlation between REF quality score and number of authors, as recorded in Scopus, for REF2021 articles published 2014–20, by UoA or Main Panel. Main panels include all articles in their UoAs, after eliminating duplicates

4 | RESULTS

4.1 | RQ1: Author counts and REF2021 quality scores

At the level of the Units of Assessment (UoAs) in which research outputs were evaluated in REF2021, there are statistically significant moderate positive correlations (0.2–0.4) between author counts and research quality in all health, life, and physical sciences (Figure 2). There are also moderate correlations (above 0.2) in two social sciences: Archaeology, and Sport, Exercise Sciences, Leisure, and Tourism (possibly party due to the inclusion of medical topics within sports research). The confidence intervals tend to be quite wide for the other fields, but most social science, arts and humanities areas have weak correlations or possibly no correlation between coauthorship and research quality. The data is consistent with some humanities areas having weak negative associations between coauthorship and research quality, but the confidence intervals for the negative correlations all include 0 (marginally for UoA 32).

When Scopus broad fields are used for the correlations instead of REF UoAs, the pattern is similar (Figure 3) but there is evidence of a negative association between coauthorship and research quality in Decision Sciences. This broad field includes three quite diverse subfields: Information Systems and Management; Management Science and Operations Research; and Statistics, Probability and Uncertainty. It is possible that the negative association is due to one of these fields tending to score higher in the REF (e.g., due to one large strong department) and tending not to coauthor. The correlation is statistically significantly positive for the Arts and

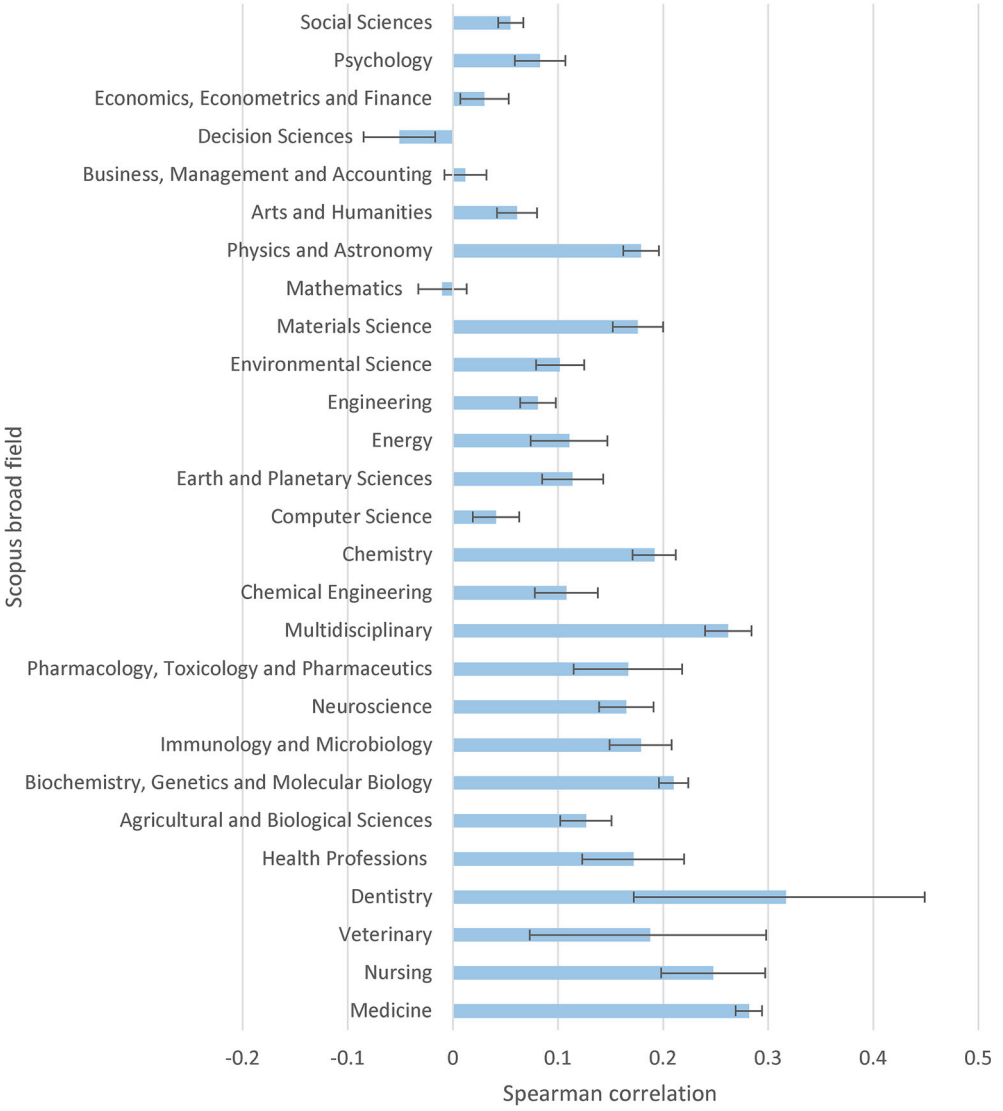

FIGURE 3 Spearman correlation between REF quality score and number of authors, as recorded in Scopus, for REF2021 articles published 2014–20, by Scopus broad field (grouped by top level field)

Humanities, but this is not strong evidence because the set may include articles from other fields that are additionally classified as Arts and Humanities (e.g., electronic music technology research). Because of these caveats, the Scopus results do not contradict the REF UoA results, and make the additional tentative suggestion of a negative association in Decision Sciences, at least as defined by Scopus.

When using the Dimensions FOR codes, the results are broadly consistent with the previous two graphs (Figure 4) and do not give additional insights. All correlations are either positive or include zero in their confidence intervals, with a mix of positive (FOR code 18) and negative (19 and 22) correlations in the arts and humanities.

Of the three classification schemes, the maximum correlation is stronger for UoAs (0.39) than for Scopus (0.31) and Dimensions (0.29). Since the UoA scheme is the finest-grained and author led, it is possibly the most accurate and hence the most reliable even though it is certainly imperfect. The higher maximum correlations support this because the norm referencing in the REF and its large size mean that combining fields would not tend to inflate correlations by grouping low scoring low authorship fields with high scoring high authorship fields because there should not be major differences between fields in scores (although there are in the arts and humanities book-based subjects because low article scores can be offset by high monograph scores). Thus, the UoA graph seems to be the most reliable, but the Scopus decision sciences negative correlation finding is an additional result.

4.2 | RQ2: Author numbers and research quality

The underlying shape of the relationship between the number of authors and the REF quality score of articles

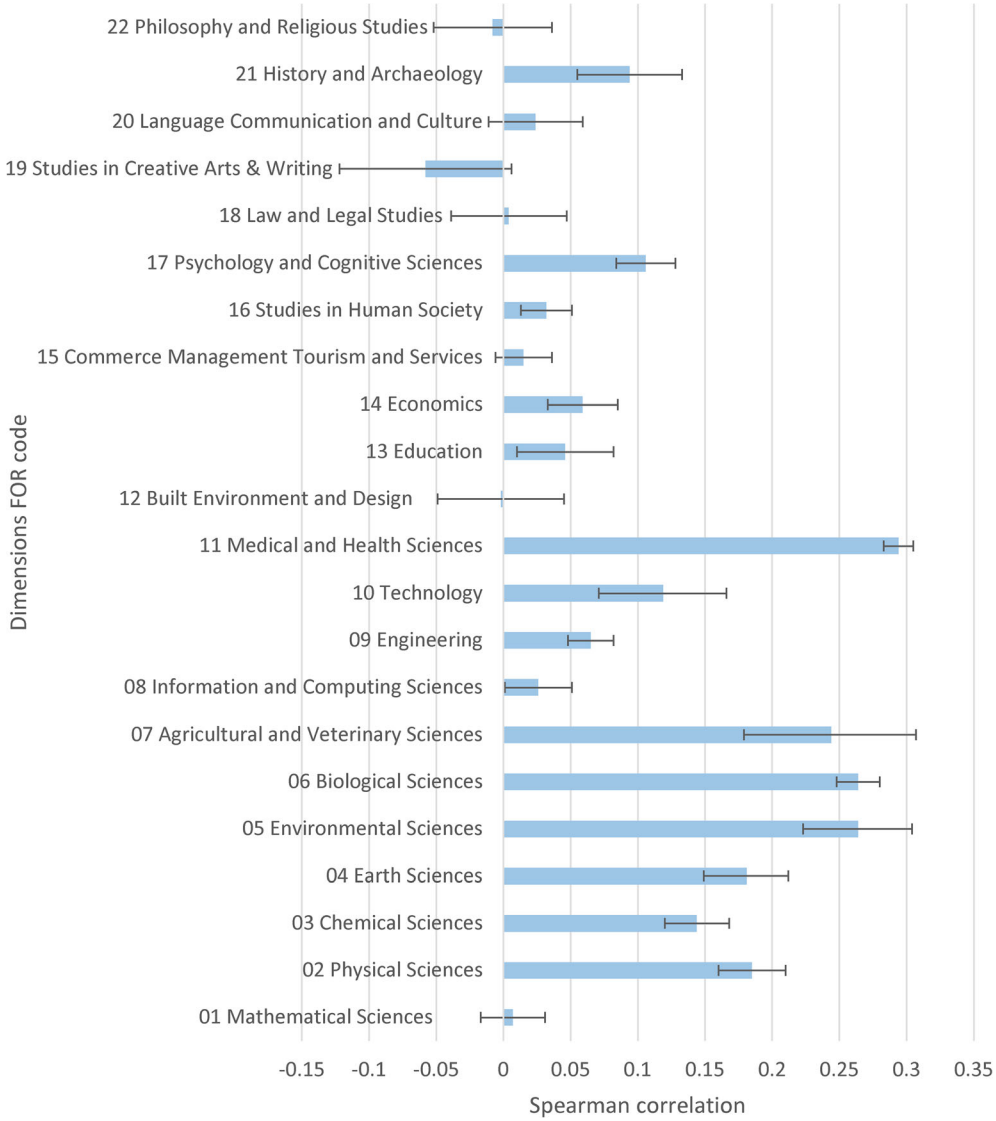

FIGURE 4 Spearman correlation between REF quality score and number of authors, as recorded in Scopus, for REF2021 articles published 2014–20, by Dimensions FOR code

is approximately logarithmic for fields with a positive correlation, as is clear for most of UoAs 1 to 12 and consistent with the rest (Figure 5). Most importantly, there is a broadly monotone increasing shape in all cases where the 95% confidence interval excludes zero except possibly UoA 12 Engineering and (although with small sample sizes) UoAs 19, 21, 23, 27 (the largest teams possibly lower quality than medium teams). With these main exceptions, the data suggests that increasing team size is an almost universal advantage in the fields for which the correlation is positive.

Uniquely, the U-shape for UoA 32 (negative correlation with 0 at the limit of the confidence interval) suggests that single author research is valued in UoA 32 Art and Design, as is common in the arts and humanities. Solo research in UoA 14 Geography (Figure 6) may also be more valued than small team research, however, although the difference is small. The UoA 32 U-shape also raises the possibility that the mixed correlation signs (positive and negative) in the humanities may therefore be side effect of the proportions of articles with the different numbers of authors.

4.3 | RQ3: Author counts and field normalized citation counts

In contrast to the situation for REF quality scores, for the same set of articles there is a statistically significant positive correlation between the number of authors and the field normalized citation score for *all* UoAs (Figure 6). This suggests that the association between authorship and citations is universal.

For one UoA (32) the 95% confidence interval for the correlation between author numbers and research quality (Figure 2) excludes positive values and the 95% confidence interval for the correlation between author numbers and citation rate (Figure 7) is exclusively positive.

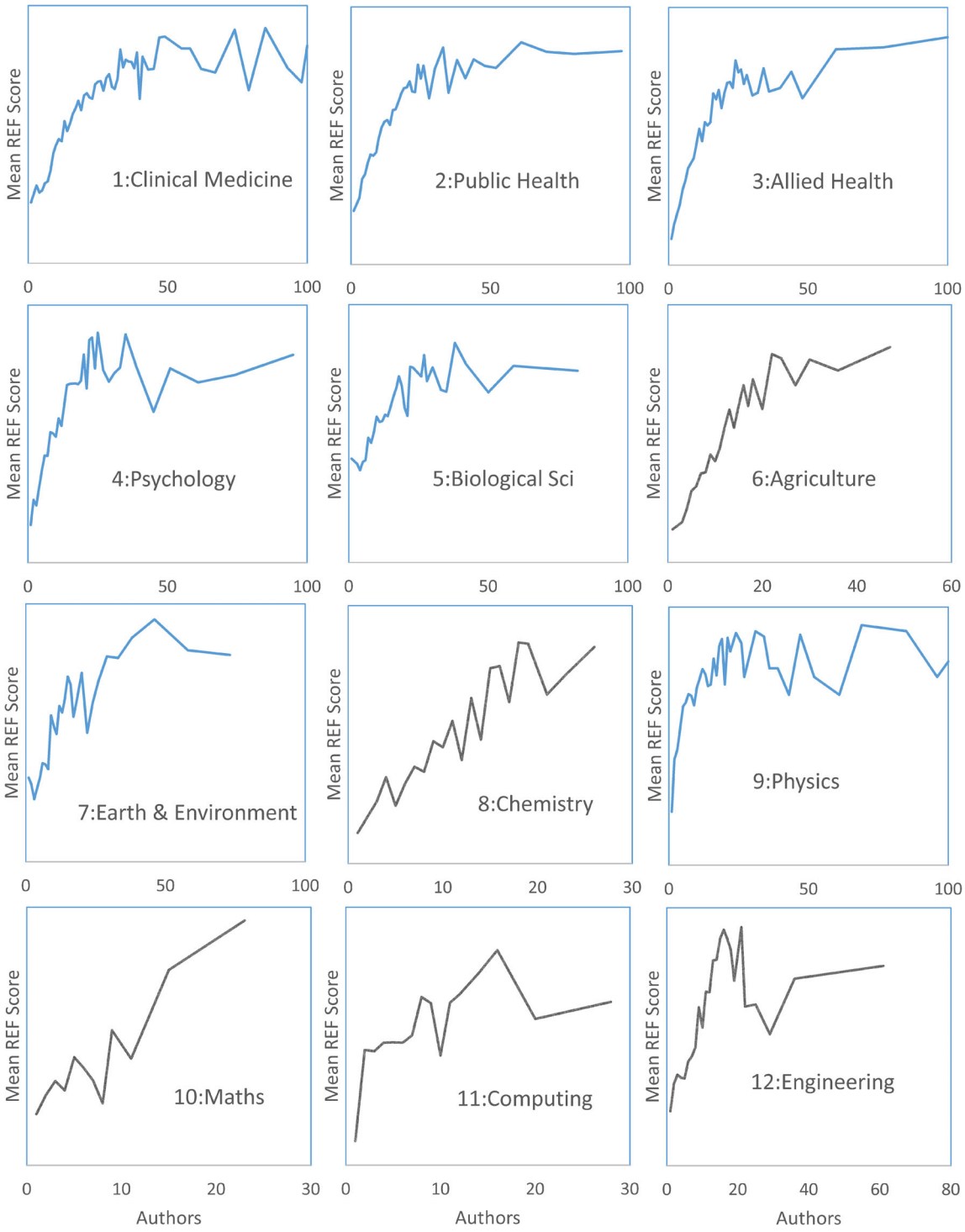

FIGURE 5 Average REF quality score against number of authors, as recorded in Scopus, for REF2021 articles published 2014–20, for UoAs 1–12. Authors are bucketed into groups of at least 25, with the right-hand value containing the all larger teams and the *x* axis value being the minimum in the group (e.g., the value at 60 for UoA 12 includes all papers with at least 60 authors). For convenience, colors indicate whether the right-hand *x* axis value is 100 or not. *Y* axis scales are redacted due to an abundance of caution with data security

Thus, for UoA 32 there is statistical evidence that increased coauthorship does not associate with increased quality but does associate with increased citations, supporting an audience effect in this case. The fact that the correlations are statistically significantly positive in all UoAs in Figure 7 and are sometimes negative in Figure 2 and are often close to zero also gives additional collective evidence that it is possible for increased citations to

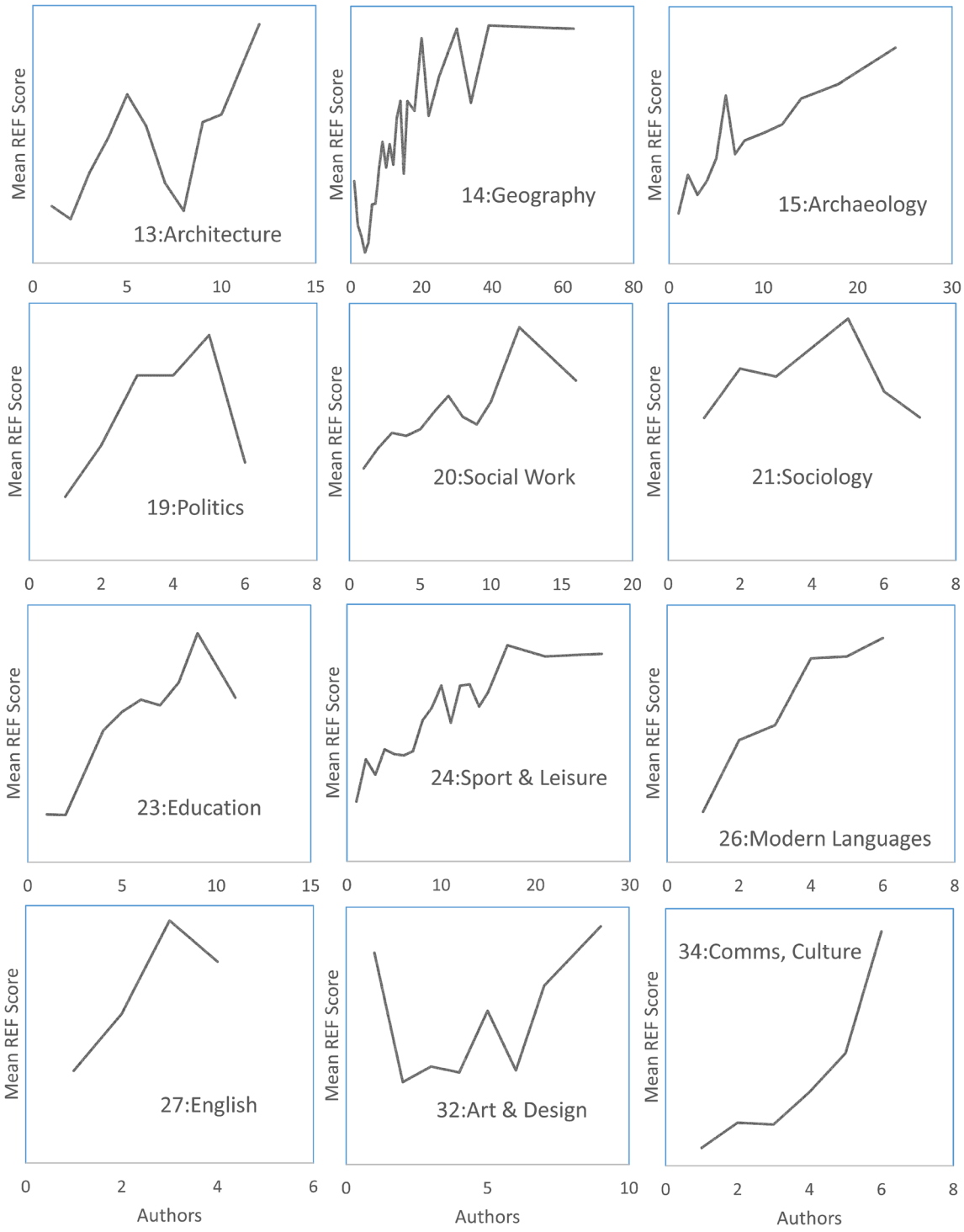

FIGURE 6 Average REF quality score against number of authors, as recorded in Scopus, for REF2021 articles published 2014–20, for UoAs 13–34 where the 95% confidence interval for the correlation excludes 0. Authors are bucketed into groups of at least 25, with the right hand value containing the all larger teams and the x axis value being the minimum in the group. *Y* axis scales are redacted due to an abundance of caution with data security

associate with more authors irrespective of article quality (i.e., possible audience effects). The situation is confirmed for Scopus broad fields (Figure 8), except in the case of Veterinary, possibly due to its small sample size.

With the Dimensions FOR codes, there are some areas of the humanities (18, 19, 22, some of 20, 21) where collaboration has quite a strong association with Scopus citation counts (Figure 9), even though there is little or

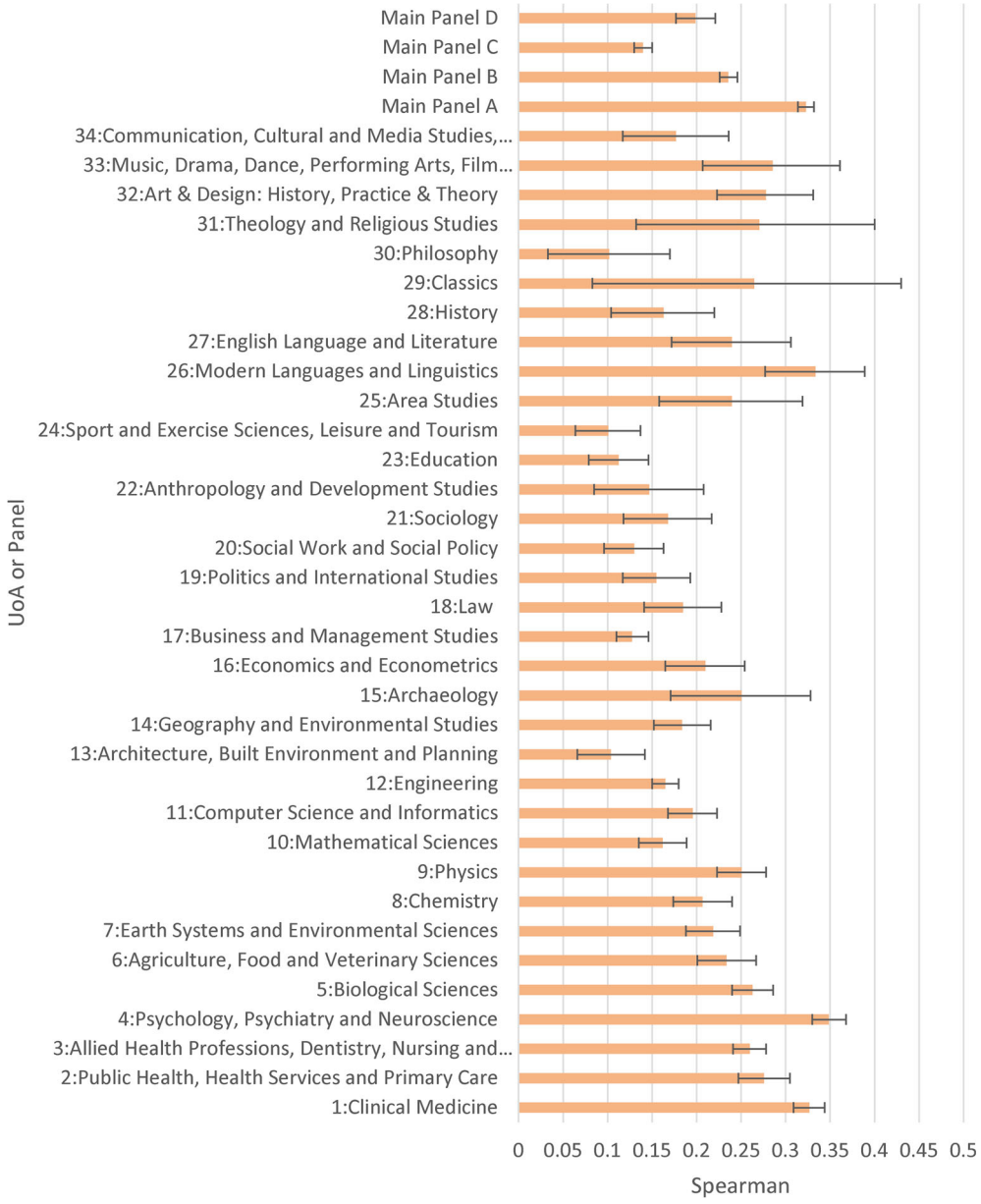

FIGURE 7 Spearman correlation between field normalized Scopus citation count and number of authors, as recorded in Scopus, for REF2021 articles published 2014–20, by UoA or Main Panel. Main panels include all articles in their UoAs, after eliminating duplicates

no association between collaboration and research quality for the same categories (Figure 5). This gives additional evidence for the audience effect from this different perspective.

The underlying shape of the relationship between the number of authors and field normalized citation counts NLCS tends to be very approximately linear for most UoAs with enough data to show a shape, but approximately logarithmic for UoA 2 (Figures 10–12). Most importantly, the shapes are broadly monotonically increasing, with the exceptions being based on little data, such as the decreasing citation rate for the largest team size in UoA 26. Monotonically increasing citation rates are consistent with the audience effect and the fact that the shapes are more linear than logarithmic is also consistent with the audience effect because each additional author potentially brings a new audience, irrespective of how many other authors there are. In contrast, the logarithmic shape suggests that beyond a certain number of authors, additional collaboration is not useful. Of course, this is a simplification because large team research might involve groups of researchers from different institutions, perhaps often including junior researchers with little "audience" of their own. The UoA 2 exception (seeming to peak in average citation rate after about 20 authors) might be due to complicating factors like this, however.

The logarithmic shapes in the quality versus authors Figure 5 compared to the linear shapes in most of Figure 10 suggest, but do not prove, that additional authors associate with the biggest improvements in research quality for smaller team sizes (e.g., adding a second author would predict a much bigger quality increase than adding a 50th author) but that there is no equivalent

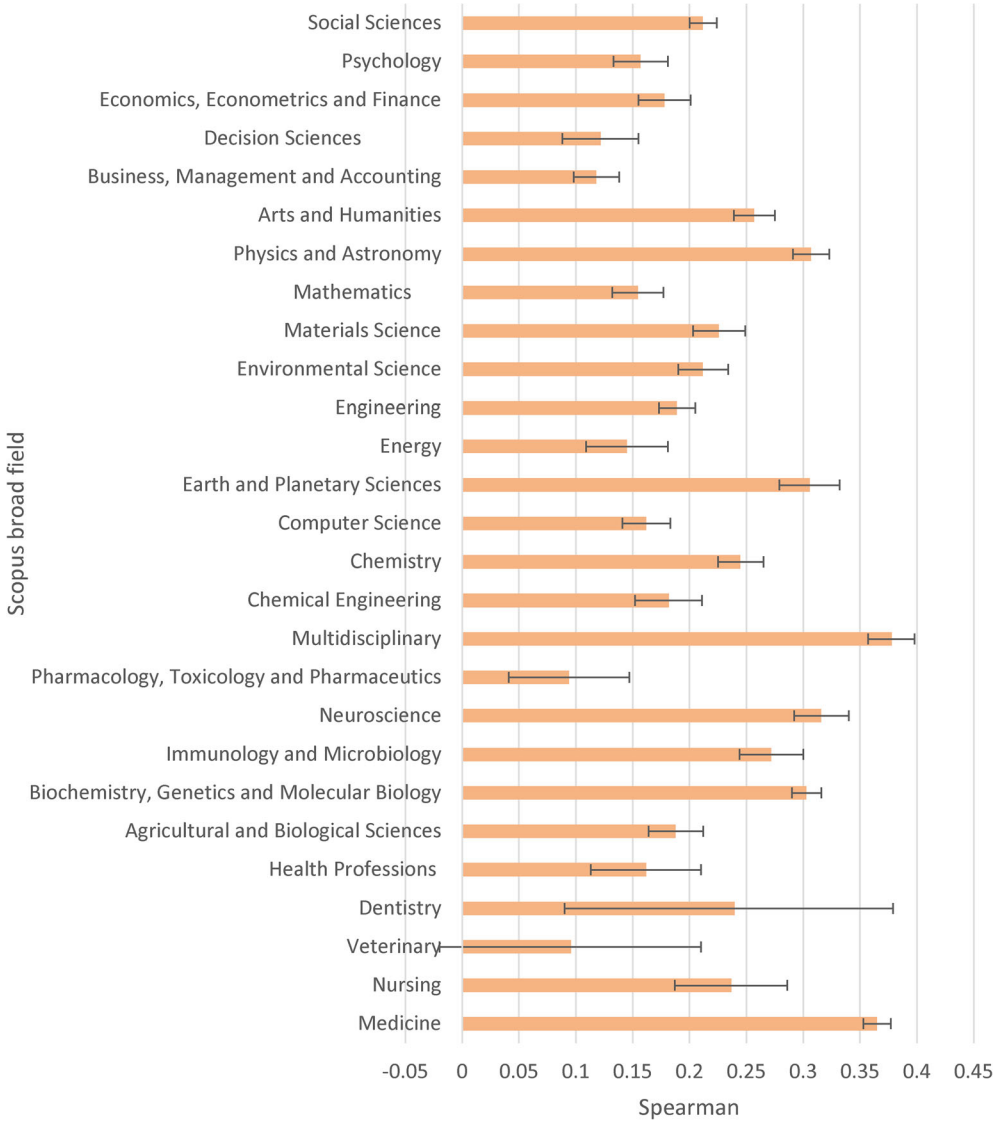

FIGURE 8 Spearman correlation between field normalized Scopus citation count and number of authors, as recorded in Scopus, for REF2021 articles published 2014–20, by Scopus broad field (grouped by top level field)

size effect for citation impact. This shape difference is therefore another source of evidence for the existence of an audience effect, at least in medicine, health, physical and life sciences.

UoA 32 Art and Design: History, Practice and Theory gives perhaps the clearest evidence that in some contexts the audience effect outweighs the quality effect. For this UoA, solo research is the least cited (Figure 12) but has a substantial quality advantage over papers with 2–6 authors (Figure 6).

5 | DISCUSSION

The results are limited by the UK focus. Although international articles are included, at least one author must have had a UK affiliation (in July 2020). Coauthorship has international differences in relationships with citations (Thelwall & Maflahi, 2020), so it is possible that different relationships between quality and author numbers might be found in some countries or for domestic research. The data is also self-selected to be the highest quality outputs of UK academics, which probably reduces the strength of the correlations found, due to relatively few articles having low scores. Similarly, the relatively coarse-grained quality scheme limits the practical possibility of obtaining high correlations. The results also reveal nothing about review articles, books, and other research outputs. It seems possible, for example, that reviewers would usually consider solo humanities monographs to be better than collaborative books.

5.1 | RQ1: Author counts and REF2021 quality scores

Our results disagree with a previous study using a different method (Bornmann, 2017) and partly agree with the similar prior study of mixed output types (Franceschet &

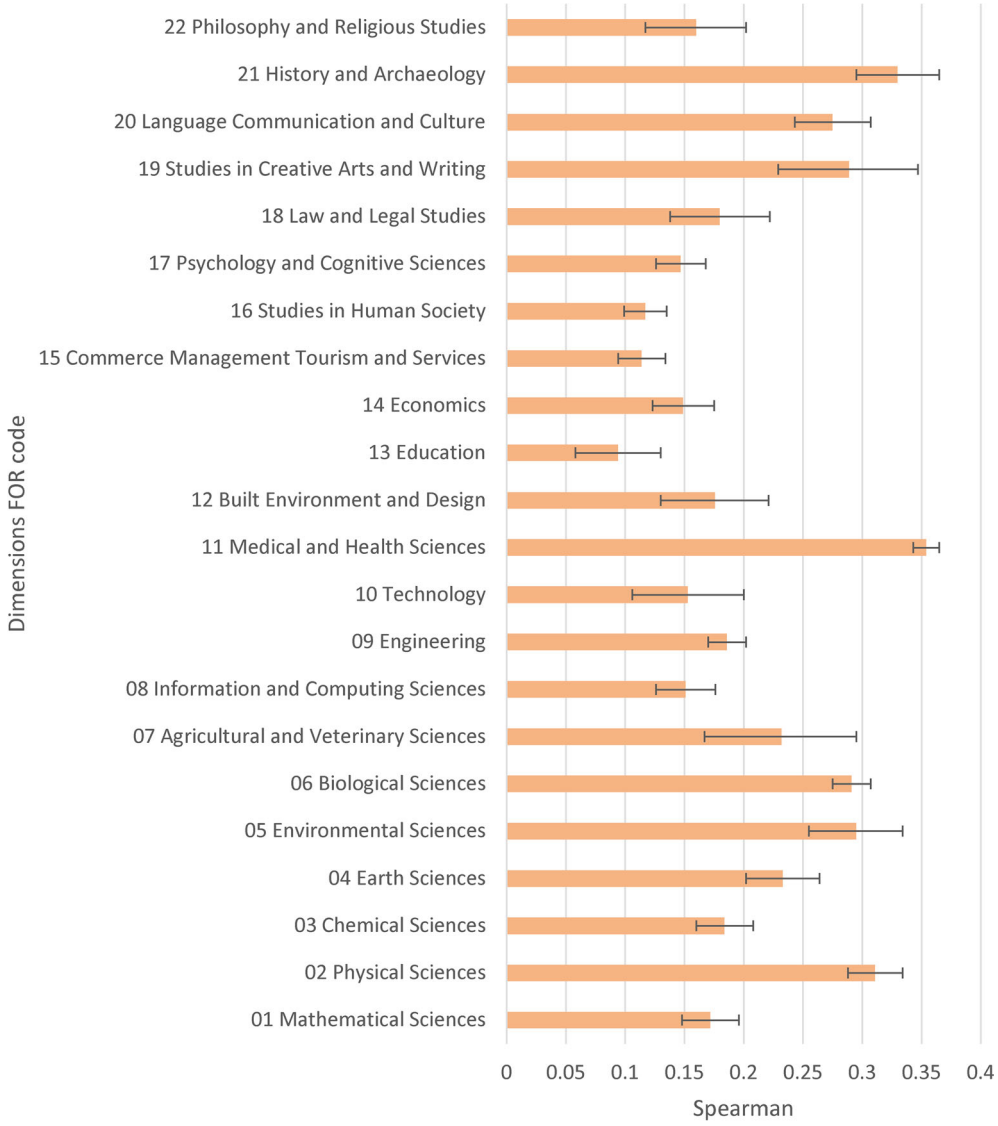

FIGURE 9 Spearman correlation between field normalized Scopus citation count and number of authors, as recorded in Scopus, for REF2021 articles published 2014–20, by Dimensions FOR code

Costantini, 2010), in terms of approximately matching UoAs to the Italian field groupings, but our results show a clear positive relationship between author numbers and research quality in areas related to the Italian categories of Chemistry, Mathematics and Computer Science (although very weak), Civil Engineering and Architecture, and Industrial and Information Engineering. Nevertheless, there was little or no relationship between author numbers and research quality for Law in both the UK and Italy (recall that the Italian data is mainly based on nonjournal outputs for Law).

The correlation results do not prove a cause-and-effect relationship in the sense that adding extra members to a team for the same research would tend to improve its quality. There are many alternative explanations, as summarized in the top half of Figure 1. For example, team leaders with a track record of high-quality research may be funded to attract more collaborators. The diversity of types of research within all the categories reported above (e.g., theoretical vs empirical; different topics; different methods), make it impossible to find a clear explanation for the relationships found.

Cause-and-effect relationships are also plausible, as summarized in the bottom half of Figure 1. For example, in empirical areas in the health domain, larger numbers of coauthors may reflect either larger-scale studies (more data collection centers) or more complex analyses (more methods), which would tend to produce more powerful findings. In contrast, in arts and humanities and some social science topics, teams may often represent mentor/mentee relationships or groups of colleagues with a common interest and similar expertise topics rather than tending to enable more powerful analyses. Since funding encourages collaboration (Davies et al., 2022), successful researchers in areas with less project-based funding may also prefer to work alone and may not need teams of people to help their research process.

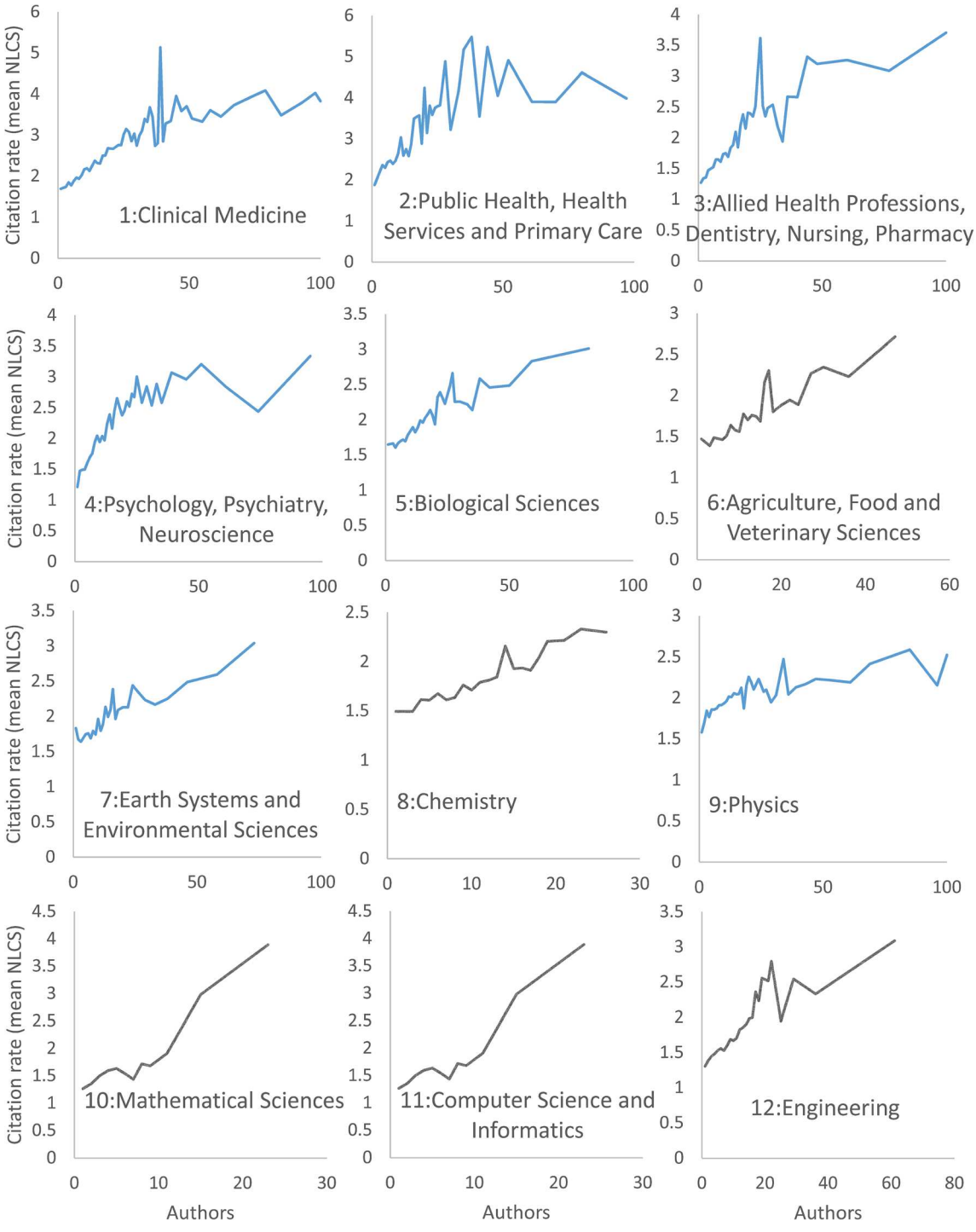

FIGURE 10 Average field and year normalized citation score (NLCS) against number of authors, as recorded in Scopus, for REF2021 articles published 2014–20, for UoA 1–12

5.2 | RQ2: Author numbers and research quality

The shape of the relationship between research quality and author numbers has not been assessed before and is therefore a new finding. Although the scale and partly the shape is dependent on the REF scoring system, the monotone increasing shape suggests that the positive correlations from RQ1 reflect a general increasing tendency rather than, for example, a U-shaped graph indicating that moderate numbers of authors are the least desirable or an inverted U shape indicating that moderate numbers of authors are the most desirable. Thus, the correlations can be given their natural interpretation.

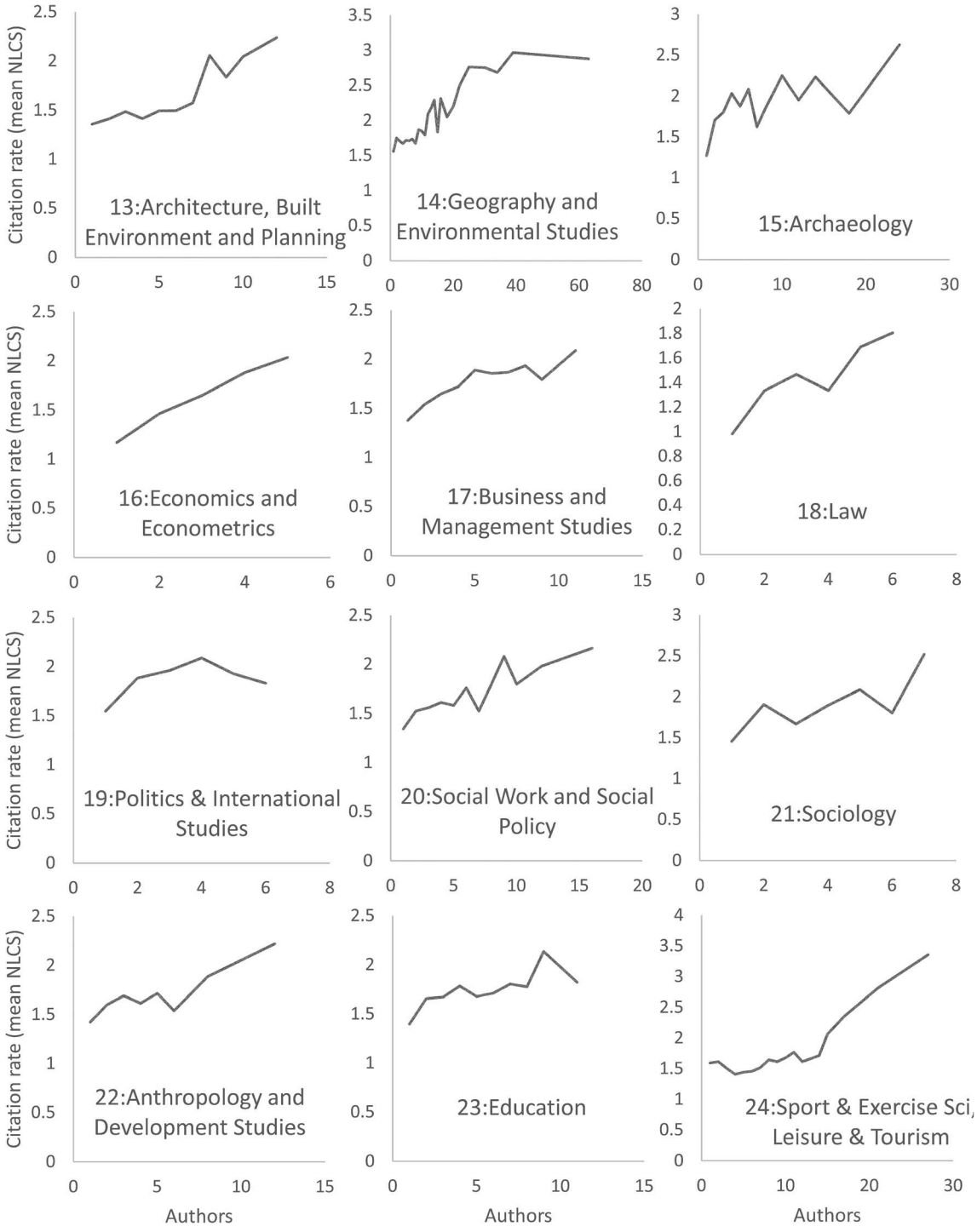

FIGURE 11 Average field and year normalized citation score (NLCS) against number of authors, as recorded in Scopus, for REF2021 articles published 2014–20, for UoA 13–24

5.3 | RQ3: Author counts and field normalized citation counts

The evidence of moderate positive correlations between author numbers and citations in some fields for which there are little, no, or negative correlations between author numbers and article quality shows that nonquality factors must be the cause of increased citations for larger team papers in some fields. This is supported by the slightly different shapes in the relationships between author numbers and citations or quality scores, even in fields for which both correlations are positive. The previously hypothesized audience effects are plausible, and the results suggest that they are more important than quality in some fields and

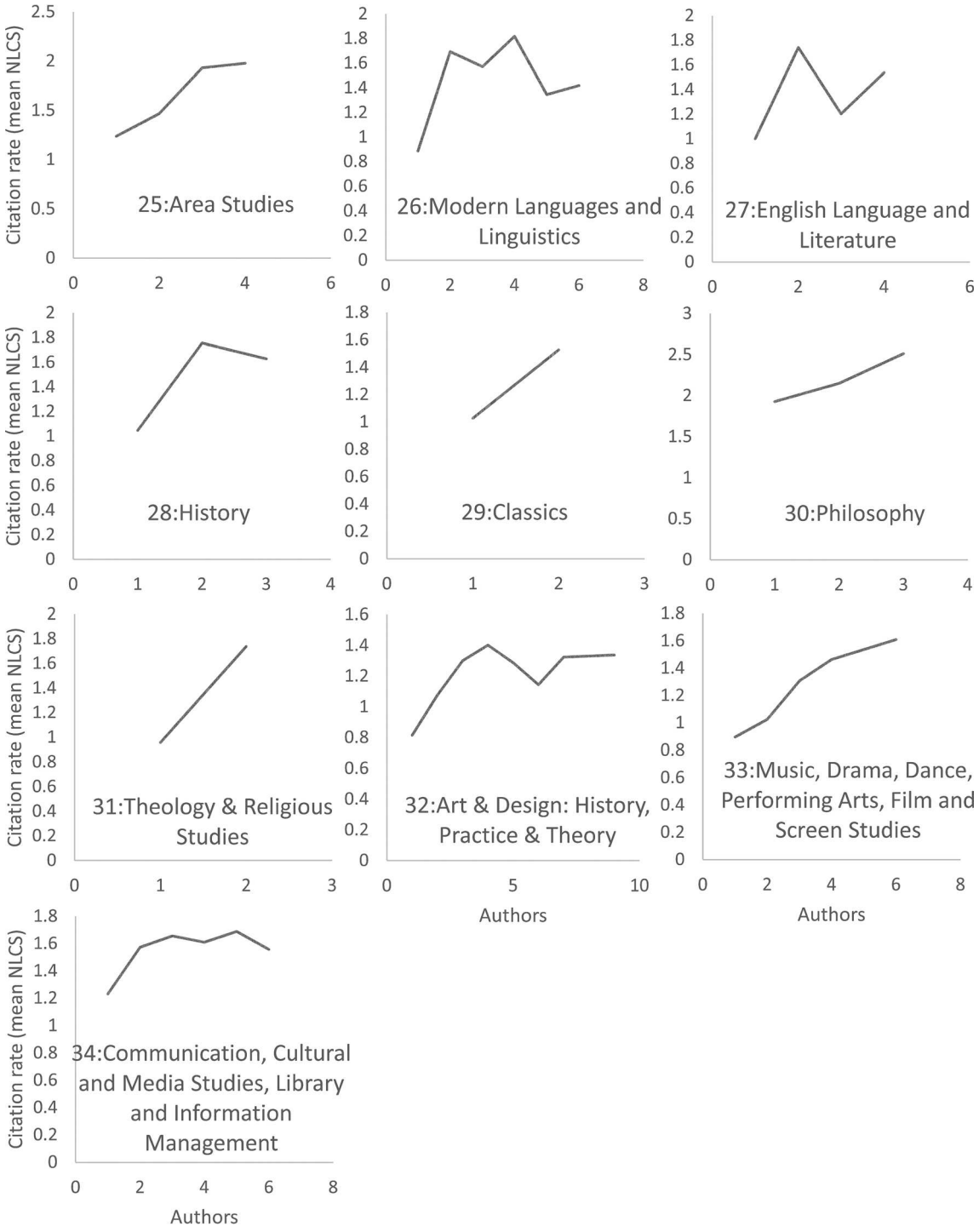

FIGURE 12 Average field and year normalized citation score (NLCS) against number of authors, as recorded in Scopus, for REF2021 articles published 2014–20, for UoA 25–34

the reverse—or possibly irrelevant—in others. Nevertheless, the results may also be due to nonaudience, nonquality factors, such as a possible tendency for authors to collaborate more on more citable topics and self-citations. In any case, in these fields where coauthorship associates with citations but not quality, coauthored research has an unfair advantage with citation-based indicators.

6 | CONCLUSION

The results show, for the first time, the strength of association between team size and research quality across all areas of science for nonreview journal articles, albeit with a UK focus. They give clear evidence of larger teams associating moderately with higher quality research, as

judged by careful consideration from REF field experts, in the health, life, and physical sciences. They show that weak positive associations are present for engineering, and many social sciences, but that there is little or no association in the arts and humanities. There is also evidence of a negative association in decision sciences. Policy makers, research funders and research managers might take these findings into account when deciding how strongly to promote collaborative research, although there are also relevant political and other considerations.

The results also show that factors other than article quality are needed to explain the tendency for articles with more authors to be more cited in some fields. This is the first statistical evidence that the audience effect (or other nonquality factors) exists independently of research quality. A possible explanation is that more authors attract greater attention to research in some fields or increased self-citation from larger teams. In the fields where this is relevant, citation-based indicators give an advantage to more coauthored research, which should be considered in research evaluations, if possible.

Of course, the results do not prove that all types of collaboration are beneficial or reveal which types are beneficial. Thus, they are not useful for individual decisions about whether to add extra scholars to a team or whether it would be beneficial to create a larger group for a study. They also do not address the issue of research productivity.

In terms of policy relevance, although the results do not prove cause-and-effect, they tend to confirm the importance of collaboration in the health, life, and physical sciences, which should reassure funders that mandate team formation or that encourage larger consortium bids. They also suggest that in engineering, the social sciences, a more cautious approach should be taken because high-quality research is almost as likely to be conducted by smaller and larger groups of coauthors. It would be a shame to provide strong collaboration incentives to researchers when there is little benefit because this might reduce research diversity and restrict the opportunities for researchers to investigate topics that do not need extensive teamwork. In the decision sciences and arts and humanities, funders and managers should think very carefully before incentivizing collaboration, however, and should avoid using citation data to support any arguments for its value.

ACKNOWLEDGMENTS

This study was funded by Research England, Scottish Funding Council, Higher Education Funding Council for Wales, and Department for the Economy, Northern Ireland as part of the Future Research Assessment Programme (https://www.jisc.ac.uk/future-research-assessment-programme). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

ORCID

Mike Thelwall https://orcid.org/0000-0001-6065-205X

REFERENCES

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.

Avkiran, N. (1997). Scientific collaboration in finance does not lead to better quality research. *Scientometrics*, *39*(2), 173–184.

Beaver, D. D. (2013). The many faces of collaboration and teamwork in scientific research: Updated reflections on scientific collaboration. *COLLNET Journal of Scientometrics and Information Management*, *7*(1), 45–54.

Benos, D. J., Bashari, E., Chaves, J. M., Gaggar, A., Kapoor, N., LaFrance, M., & Zotov, A. (2007). The ups and downs of peer review. *Advances in Physiology Education*, *31*, 145–152.

Bornmann, L. (2017). Is collaboration among scientists related to the citation impact of papers because their quality increases with collaboration? An analysis based on data from F1000Prime and normalized citation scores. *Journal of the Association for Information Science and Technology*, *68*(4), 1036–1047.

Cantor, R. M., Lange, K., & Sinsheimer, J. S. (2010). Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *American Journal of Human Genetics*, *86*(1), 6–22.

D'ippolito, B., & Rüling, C. C. (2019). Research collaboration in large scale research infrastructures: Collaboration types and policy implications. *Research Policy*, *48*(5), 1282–1296.

Davies, B., Gush, J., Hendy, S. C., & Jaffe, A. B. (2022). Research funding and collaboration. *Research Policy*, *51*(2), 104421.

de Solla Price, D. J. (1963). Little science, big science. In *Little science, big science*. Columbia University Press.

Deichmann, D., Moser, C., Birkholz, J. M., Nerghes, A., Groenewegen, P., & Wang, S. (2020). Ideas with impact: How connectivity shapes idea diffusion. *Research Policy*, *49*(1), 103881.

Didegah, F., & Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, *7*(4), 861–873.

Dimensions.ai. (2021). *Which research categories and classification schemes are available in Dimensions?* Retrieved from https://plus.dimensions.ai/support/solutions/articles/23000018820-which-research-categories-and-classification-schemes-are-available-in-dimensions

Ding, J., Shen, Z., Ahlgren, P., Jeppsson, T., Minguillo, D., & Lyhagen, J. (2021). The link between ethnic diversity and scientific impact: The mediating effect of novelty and audience diversity. *Scientometrics*, *126*(9), 7759–7810.

Finlay, S. C., Ni, C., & Sugimoto, C. R. (2012). New methods for an old debate: Utilizing reader response to investigate the relationship between collaboration and quality in academic journal articles. *Library & Information Science Research*, *34*(2), 131–137.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, *10*(4), 507–521.

Franceschet, M., & Costantini, A. (2010). The effect of scholar collaboration on impact and quality of academic papers. *Journal of Informetrics*, *4*(4), 540–553.

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*. Sage.

Gilbert, S., & Green, C. (2021). *Vaxxers: The inside story of the Oxford AstraZeneca vaccine and the race against the virus*. Hachette UK.

Glänzel, W., & Thijs, B. (2004). Does co-authorship inflate the share of self-citations? *Scientometrics*, *61*(3), 395–404.

Gülen, S., Fonnes, S., Andresen, K., & Rosenberg, J. (2020). More than one-third of Cochrane reviews had gift authors, whereas ghost authorship was rare. *Journal of Clinical Epidemiology*, *128*, 13–19.

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), 532–548.

Hansson, J., Golub, K., Tyrkkö, J., & Ahlström, I. (2021). Publication practices in the humanities: An in-depth study of a Swedish arts and humanities faculty 2010-2018. *Nordic Journal of Library and Information Studies*, *2*(2), 41–64.

He, Z. L., Geng, X. S., & Campbell-Hunt, C. (2009). Research collaboration and research output: A longitudinal study of 65 biomedical scientists in a New Zealand university. *Research Policy*, *38*(2), 306–317.

Hook, D. W., Porter, S. J., & Herzog, C. (2018). Dimensions: Building context for search and evaluation. *Frontiers in Research Metrics and Analytics*, *3*, 23.

Hull, D. L. (2010). *Science as a process*. University of Chicago Press.

Ignaciuk, A., Rice, M., Bogardi, J., Canadell, J. G., Dhakal, S., Ingram, J., & Rosenberg, M. (2012). Responding to complex societal challenges: A decade of earth system science partnership (ESSP) interdisciplinary research. *Current Opinion in Environmental Sustainability*, *4*(1), 147–158.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18.

Lancho Barrantes, B. S., Guerrero Bote, V. P., Rodríguez, Z. C., & de Moya Anegón, F. (2012). Citation flows in the zones of influence of scientific collaborations. *Journal of the American Society for Information Science and Technology*, *63*(3), 481–489.

Langfeldt, L., Nedeva, M., Sörlin, S., & Thomas, D. A. (2020). Co-existing notions of research quality: A framework to study context-specific understandings of good research. *Minerva*, *58*(1), 115–137.

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.

Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, *35*(5), 673–702.

Lin, W. Y. C., & Huang, M. H. (2012). The relationship between co-authorship, currency of references and author self-citations. *Scientometrics*, *90*(2), 343–360.

Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology, & Human Values*, *17*(1), 101–126.

Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & Delgado López-Cózar, E. (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.

Melin, G. (2000). Pragmatism and self-organization: Research collaboration on the individual level. *Research Policy*, *29*(1), 31–40.

Perkins, D. H. (2000). *Introduction to high energy physics*. Cambridge University Press.

Presser, S. (1980). Collaboration and the quality of research. *Social Studies of Science*, *10*(1), 95–101.

Real, L. A. (2012). Collaboration in the sciences and the humanities: A comparative phenomenology. *Arts and Humanities in Higher Education*, *11*(3), 250–261.

REF2021. (2020). *Panel criteria and working methods*. Retrieved from https://www.ref.ac.uk/media/1450/ref-2019_02-panel-criteria-and-working-methods.pdf

Schmoch, U., & Schubert, T. (2008). Are international co-publications an indicator for quality of scientific research? *Scientometrics*, *74*(3), 361–377.

Shaw, P., Phillips, A., & Gutiérrez, M. B. (2022). The death of the monograph? *Publishing Research Quarterly*, *38*, 382–395.

Shen, H., Xie, J., Li, J., & Cheng, Y. (2021). The correlation between scientific collaboration and citation count at the paper level: A meta-analysis. *Scientometrics*, *126*(4), 3443–3470.

Shen, W. Q., Liu, D., & Chen, H. (2017). Chinese Ph.D. students on exchange in European Union countries: Experiences and benefits. *European Journal of Higher Education*, *7*(3), 322–335.

Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology*, *41*(1), 643–681.

Sooryamoorthy, R. (2009). Do types of collaboration change citation? Collaboration and citation patterns of South African science publications. *Scientometrics*, *81*(1), 177–193.

Technopolis. (2015). *REF accountability review: Costs, benefits and burden*. Retrieved from https://www.technopolis-group.com/report/ref-accountability-review-costs-benefits-and-burden/

Thelwall, M. (2016). Interpreting correlations between citation counts and other indicators. *Scientometrics*, *108*(1), 337–347.

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. (2020). 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., & Maflahi, N. (2015). Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers. *Journal of the Association for Information Science*

and Technology, *66*(6), 1124–1135. https://doi.org/10.1002/asi.23252

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.

Thelwall, M., & Sud, P. (2016). National, disciplinary and temporal variations in the extent to which articles with more authors have more impact: Evidence from a geometric field normalised citation indicator. *Journal of Informetrics*, *10*(1), 48–61. https://doi.org/10.1016/j.joi.2015.11.007

Ubfal, D., & Maffioli, A. (2011). The impact of funding on research collaboration: Evidence from a developing country. *Research Policy*, *40*(9), 1269–1279.

Van Rijnsoever, F. J., & Hessels, L. K. (2011). Factors associated with disciplinary and interdisciplinary research collaboration. *Research Policy*, *40*(3), 463–472.

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.

Yan, E., Wu, C., & Song, M. (2018). The funding factor: A cross-disciplinary examination of the association between research funding and citation impact. *Scientometrics*, *115*(1), 369–384.

Ziman, J. (2000). *Real science: What it is and what it means*. Cambridge University Press.

How to cite this article: Thelwall, M., Kousha, K., Abdoli, M., Stuart, E., Makita, M., Wilson, P., & Levitt, J. (2023). Why are coauthored academic articles more cited: Higher quality or larger audience? *Journal of the Association for Information Science and Technology*, *74*(7), 791–810. https://doi.org/10.1002/asi.24755