UAB
Universitat Autònoma
de Barcelona

**Dipòsit digital
de documents
de la UAB**

The strain on scientific publishing

Mark A. Hanson[1], Pablo Gómez Barreiro[2], Paolo Crosetto[3], Dan Brockington[4]

Author correspondence:
MAH (m.hanson@exeter.ac.uk, ORCID: https://orcid.org/0000-0002-6125-3672)
PGB (p.gomez@kew.org, ORCID: https://orcid.org/0000-0002-3140-3326)
PC (paolo.crosetto@inrae.fr, ORCID: https://orcid.org/0000-0002-9153-0159)
DB (Daniel.Brockington@uab.cat, ORCID: https://orcid.org/0000-0001-5692-0154)

1. *Centre for Ecology and Conservation, Faculty of Environment, Science and Economy, University of Exeter, Penryn, TR10 9FE, United Kingdom*
2. *Royal Botanic Gardens, Kew, Wakehurst, Ardingly, West Sussex RH17 6TN, United Kingdom*
3. *Univ. Grenoble Alpes, INRAE, CNRS, Grenoble INP, GAEL, Grenoble 38000, France*
4. *Institut de Ciència i Tecnologia Ambientals (ICTA), Universitat Autònoma de Barcelona & ICREA, Pg. Lluís Companys 23, Barcelona, Spain*

Abstract

Scientists are increasingly overwhelmed by the volume of articles being published. Total articles indexed in Scopus and Web of Science have grown exponentially in recent years; in 2022 the article total was ~47% higher than in 2016, which has outpaced the limited growth – if any – in the number of practising scientists. Thus, publication workload per scientist (writing, reviewing, editing) has increased dramatically. We define this problem as "the strain on scientific publishing." To analyse this strain, we present five data-driven metrics showing publisher growth, processing times, and citation behaviours. We draw these data from web scrapes, requests for data from publishers, and material that is freely available through publisher websites. Our findings are based on millions of papers produced by leading academic publishers. We find specific groups have disproportionately grown in their articles published per year, contributing to this strain. Some publishers enabled this growth by adopting a strategy of hosting "special issues," which publish articles with reduced turnaround times. Given pressures on researchers to "publish or perish" to be competitive for funding applications, this strain was likely amplified by these offers to publish more articles. We also observed widespread year-over-year inflation of journal impact factors coinciding with this strain, which risks confusing quality signals. Such exponential growth cannot be sustained. The metrics we define here should enable this evolving conversation to reach actionable solutions to address the strain on scientific publishing.

Introduction

Academic publishing has a problem. The last few years have seen an exponential growth in the number of peer-reviewed journal articles, which has not been matched by the training of new researchers who can vet those articles (Fig. 1A). Editors are reporting difficulties in recruiting qualified peer reviewers (1, 2), and scientists are overwhelmed by the immense total of new articles being published (3, 4). We will call this problem "the strain on scientific publishing."

Part of this growth may come from inclusivity initiatives or investment in the Global South, which make publishing accessible to more researchers (5, 6). Parallel efforts have also appeared in recent years to combat systemic biases in scientific publishing (7–9), including positive-result bias (10). If this strain on scientific publishing comes from such initiatives, it would be welcome and should be accommodated.

However, this strain may compromise the ability of scientists to be rigorous when vetting information (11). If scientific rigour is allowed to slip, it devalues the term "science" (12). Recent controversies already demonstrate this threat, as research paper mills operating within publishing groups have caused mass article retractions (13–15), alongside renewed calls to address so-called "predatory publishing" (16).

To understand the forces that contribute to this strain, we first present a simple schematic to describe scientific publishing. We then specifically analyse publishers, as their infrastructures regulate the rate at which growth in published articles can occur. To do this, we identify five key metrics that help us to understand the constitution and origins of this strain: growth in total articles and special issues, differences in article turnaround times or rejection rates, and a new metric informing on journal quality that we call "impact inflation."

These metrics should be viewed in light of publisher business models. First, there is the more classic subscription-based model generating revenue from readers. Second, there is the "gold open access" model, which generates revenue through article processing charges that authors pay instead. In both cases publishers can act either as for-profit or not-for-profit organisations. We therefore consider if aspects of either of these business models are contributing to the strain.

Here we provide a comparative analysis, combining multiple metrics, to reveal what has generated the strain on scientific publishing. We find strain is not strictly tied to any one publisher business model, although some behaviours are associated with specific gold open access publishers. We argue that existing efforts to address this strain are insufficient. We highlight specific areas needing transparency, and actions that publishers, researchers, and funders can take to respond to this strain. Our study provides the essential data to inform the existing conversation on academic publishing practices.

Framework and Methods

The love triangle of scientific publishing: a conceptual framework

The strain on scientific publishing is the result of interactions between three sets of players: publishers, researchers, and funders.

Publishers want to publish as many papers as possible, subject to a quality constraint. They give researchers "publication", i.e. a "badge of quality" that researchers use for their own goals. The quality of a badge is often determined by journal-level prestige metrics, such as the Clarivate journal Impact Factor (IF), or Scopus Scimago Journal Rank (SJR) (17, 18), and ultimately by association with the quality of published papers. Publishers compete with each other to attract the most and/or the best papers.

Researchers want to publish as many papers in prestigious journals as possible, subject to an effort constraint. They do so because publications and citations are key to employment, promotion, and funding: so called "publish or perish" (12, 19). Researchers act as authors that generate articles, but can also be referees and editors that consult for publishers and funders for free. In exchange, they gain influence over administering publisher badges of quality and who gets limited jobs or funding. More altruistically, they help ensure the quality of science in their field.

Funders (e.g. universities, funding agencies) use "badges" from the science publication market as measures of quality to guide their decisions on whom to hire and fund (20, 21); in some countries, journal badges directly determine promotion or salary (e.g. (22)). Ultimately, money from funders supports the whole market, and funders want cost-effective and informative signals to help guide their decisions.

The incentives for publishers and researchers to increase their output drives growth. This is not problematic per se, but it should not come at the expense of research quality. The difficulty is that "quality" is hard to define (17, 18, 23), and some metrics are at risk of abuse per ***Goodhart's law***: *"when a measure becomes a target, it ceases to be a good measure"* (24). For instance, having many citations can indicate an author, article, or journal, is having an impact. But citations can be gamed through self-citing or coordinated "citation cartels" (25, 26).

Collectively, the push and pull by the motivations of these players defines the sum product of the scientific publishing industry.

Data collection and analysis

A full summary of our data methodology is given in the supplementary materials and methods. **In brief:** we produced five metrics of publisher practice that describe the total volume of material being published, or that affect the quality of publisher "badges". We focused our analyses on the last decade of publication growth, with special attention paid to the period of 2016-2022, as pre-2016, some data types were less available. We used the Scopus database (via Scimago (27)) filtered for journals indexed in both Scopus and Web of Science. We further

assembled journal/article data by scraping information in the public domain from web pages, and/or following direct request to publishers. These metrics are:

- Total articles indexed in both Scopus and Web of Science
- Share of articles appearing in special issues
- Article turnaround times from submission to acceptance
- Journal rejection rates as defined by publishers
- A new metric we call "impact inflation," informed by journal citation behaviours

Due to limits in web scraping data availability, for special issue proportions, turnaround times, and rejection rates, we focused on only a subset of publishers and articles (Table 1). Further, due to copyright concerns over our web scraping of information in the public domain, we have been legally advised to forego a public release of our data and scripts at this time, but will make these available for formal peer review. High resolution versions of the figures can be found at doi: 10.6084/m9.figshare.24203790.

| Publisher	Web scraped journals included	Articles with turnaround times[1]	Total journals (Scimago)	Total articles (Scimago)[1]
BMC*	289	344501	213	241493
Elsevier	376	531580	1579	2988422
Frontiers	44	291017	49	329370
Hindawi*	220	226612	161	155396
MDPI	98	838448	152	840518
Nature*	144	360855	111	346845
PLOS	12	243398	7	148404
Springer	1259	1371405	1582	1372625
Taylor & Francis	1063	512438	1160	525029
Wiley	1257	890174	1432	1416024

[1] Time period 2016-2022

Table 1: summary of web scraped data informing share of special issue articles and turnaround times. For some publishers, the number of web scraped journals or articles with turnaround time data exceeds the totals from our Scimago dataset (noted with *). This is because, in this second dataset, we included all journals by a given publisher, even if they were not indexed, or indexed by only one of Scopus or Web of Science.

Results

A few publishers disproportionately contribute to total article growth

There were ~896k more indexed articles per year in 2022 (~2.82m articles) compared to 2016 (~1.92m articles) (Fig. 1A), a year-on-year growth of ~5.6% over this time period. To understand the source of this substantial growth, we first divided article output across publishers per Scopus publisher labels (Fig. 1B). The five largest publishers by total article output include Elsevier, Multidisciplinary Publishing Institute (MDPI), Wiley-Blackwell (Wiley), Springer, and Frontiers Media (Frontiers) respectively. However, in terms of strain added since
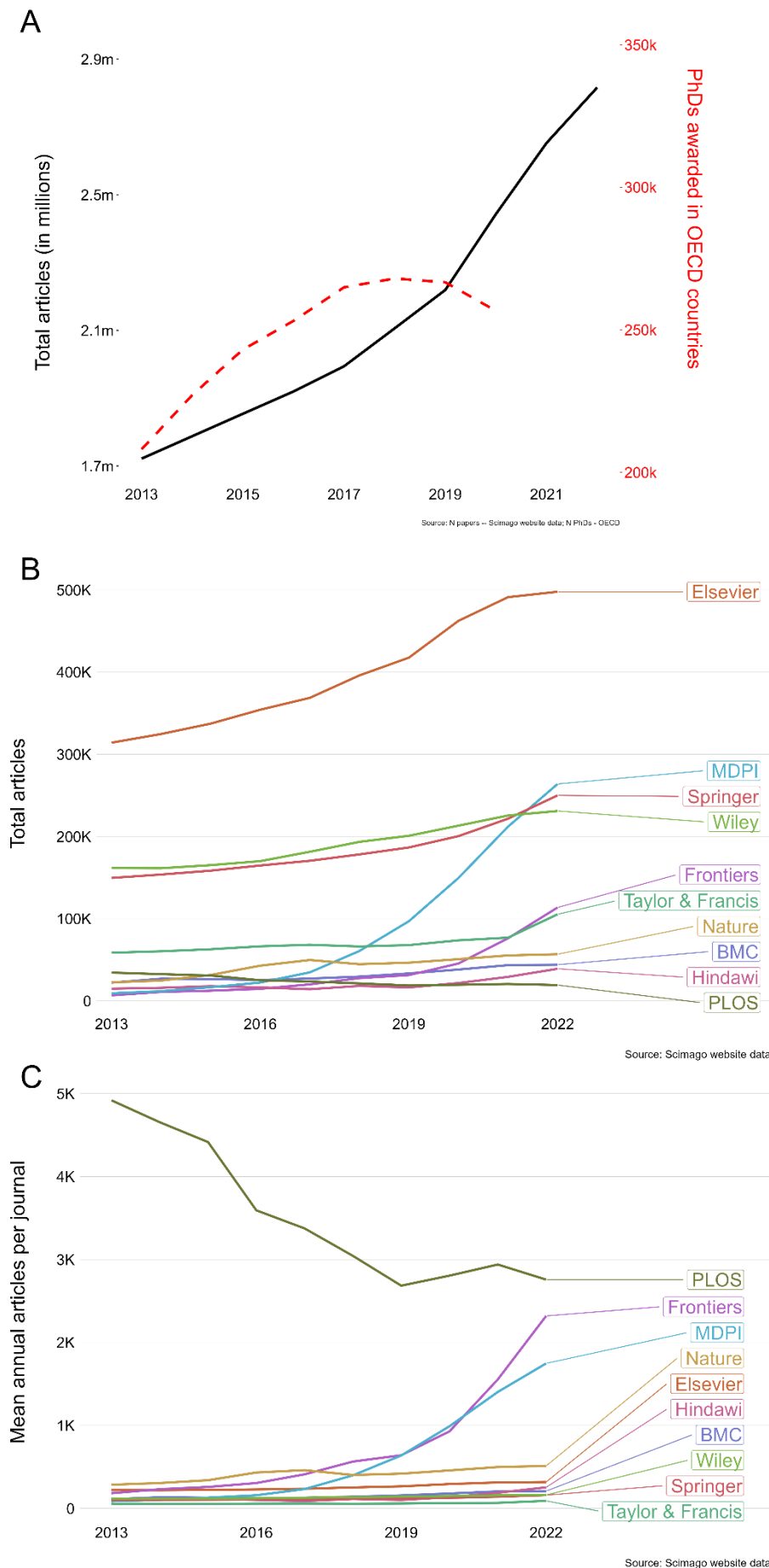
Figure 1: Total article output is increasing. A) Total articles being published per year has increased exponentially, while PhDs being awarded have not kept up. This remains true with addition of non-OECD countries, or when using global total employed researcher-hours instead of PhD graduates as a proxy for active researchers (Fig. 1supp1). B-C) Total articles per year by publisher (B), or per journal per year by publisher (C). Also see growth in journals per publisher (Fig. 1supp2) and by size class (Fig. 1supp3).

Acronyms

Impact Factor (IF), Scimago Journal Rank (SJR)

Number of papers published in regular vs special issues, 2016-22
One square = 800 articles
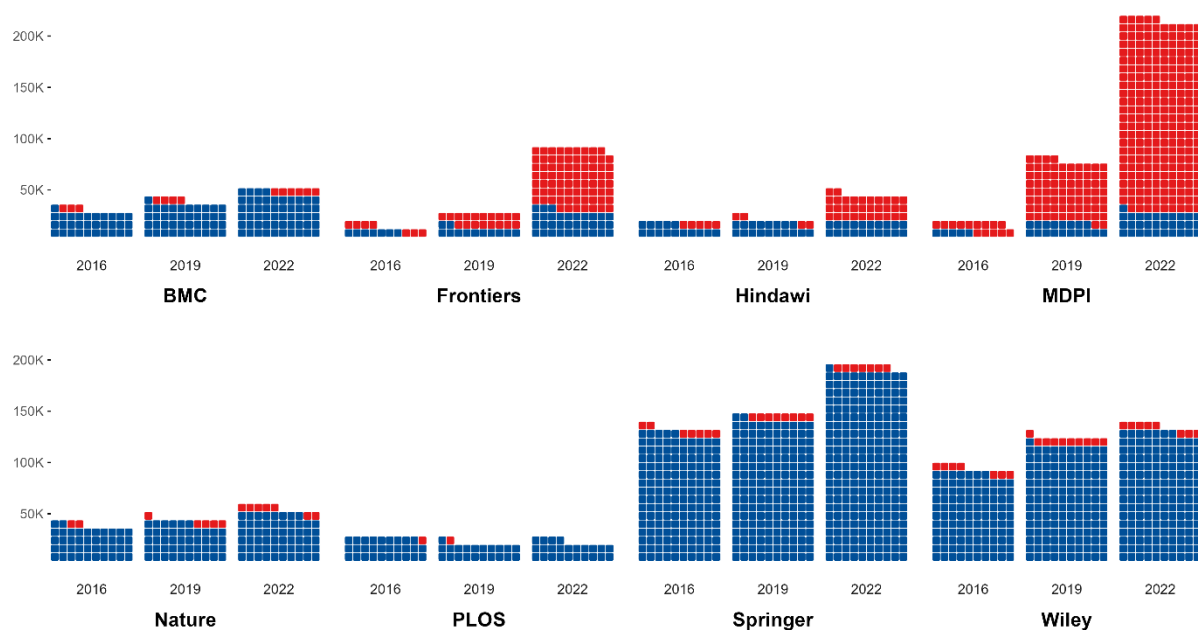
Source: data scraped from the publisher's website
Note: Special issues are called Collections at PLOS and Topics at Frontiers. For MDPI Collections, Sections and Topics not shown.

Figure 2: rise of the special issue model of publishing. Normal articles (blue) and special issue articles (red) over time. Frontiers, Hindawi, and especially MDPI publish a majority of their articles through special issues, including an increase in recent years alongside growth seen in Fig. 1 (detailed further in Fig. 2supp1,2). These data reflect only a fraction of total articles shown in Fig. 1, limited due to sampling methodology (Table 1).

2016, their rank order changes: journals from MDPI (~27%), Elsevier (~16%), Frontiers (~11%), Springer (~9.5%), and Wiley (~6.8%) have contributed >70% of the increase in articles per year. Elsevier and Springer own a huge proportion of total journals, a number that has also increased over the past decade (Figure 1supp2). As such, we normalised article output per journal to decouple the immensity of groups like Elsevier and Springer from the growth of articles itself. While Elsevier has increased article outputs per journal slightly, other groups such as MDPI and Frontiers, have become disproportionately high producers of published articles per journal (Fig. 1C).

Taken together, groups like Elsevier and Springer have quantitatively increased total article output by distributing articles across an increasing number of journals. Meanwhile groups like MDPI and Frontiers have been exponentially increasing the number of publications handled by a much smaller pool of journals. These publishers reflect two different mechanisms that have promoted the exponential increase in total articles published over the last few years.

Growth in articles published through "special issues"

"Special issues" are distinct from standard articles because they are invited by journals or editors, rather than submitted independently by authors. They also delegate responsibilities to guest editors, whereas editors for normal issues are formal staff of the publisher. In recent years, certain publishers have adopted this business model as a route to publish the majority of their articles (Fig. 2). This behaviour encourages researchers to generate articles specifically for special issues, raising concerns that publishers could abuse this model for profit

(28). Here we describe this growth in special issues for eight publishers for which we could collect data.

Between 2016 and 2022, the proportion of special issue articles grew drastically for Hindawi, Frontiers, and MDPI (Fig. 2supp1,2). These publishers depend on article processing charges for their revenues, which are paid by authors to secure gold open access licences. But this special issue growth is not a necessary feature of open access publishing as similar changes were not seen in other gold open access publishers (i.e. BMC, PLOS). Publishers using both subscription and open access approaches (Nature, Springer, Wiley) also tended to publish small proportions of special issues.

These data show that the strain generated by special issues is not a direct consequence of the rise of open access publishing per se, or associated article processing charges. Instead, the dominance of special issues in a publisher's business model is publisher-dependent.

Decreasing mean, increasing homogeneity of turnaround times

We define article turnaround times as the time taken from submission to acceptance. The peer review process can take weeks to months depending on field of research and the magnitude or type of revisions required, meaning turnaround times across articles and journals are expected to vary. Turnaround times also reflect a trade-off between rigour and efficiency: longer timeframes can allow greater rigour, but they delay publication. Shorter timeframes could reflect greater efficiency, but rushing of timeframes could make mistakes more likely. Given these considerations, there should be an objective, reasonable, minimum and maximum turnaround time needed to conduct appropriate peer review. Moreover, a journal performing rigorous peer review should have heterogenous turnaround times if each article is considered and addressed according to its unique needs.

We analysed turnaround times between 2016 and 2022 for publications where data were available. We found that average turnaround times vary markedly across publishers. Like others (29, 30), we found that MDPI had an average turnaround time of ~37 days from first submission to acceptance in 2022, a level they have held at since ~2018. This turnaround time is far lower than comparable publishers like Frontiers (72 days) and Hindawi (83 days), which also saw a decline in mean turnaround time between 2020 and 2022. On the other hand, other publishers in our dataset had turnaround times of >130 days, and if anything, their turnaround times increased slightly between 2016-2022 (Fig. 3A).

The publishers decreasing their turnaround times also show declining variances. Turnaround times for Hindawi, Frontiers, and especially MDPI are becoming increasingly homogenous (Fig. 3B and 3supp1). This implies these articles, regardless of initial quality or field of research, and despite the expectation of heterogeneity, are all accepted in an increasingly similar timeframe.

The decrease in mean turnaround times (Fig. 3A) also aligns with inflection points for the exponential growth of articles published as part of special issues in Hindawi (2020), Frontiers (2019), and MDPI (2016) (see Fig. 2supp1). We therefore asked if special issue articles are processed more rapidly than normal articles in general. For most publishers, this was indeed the case, even independent of proportions of normal and special issue articles (Fig. 3supp2).
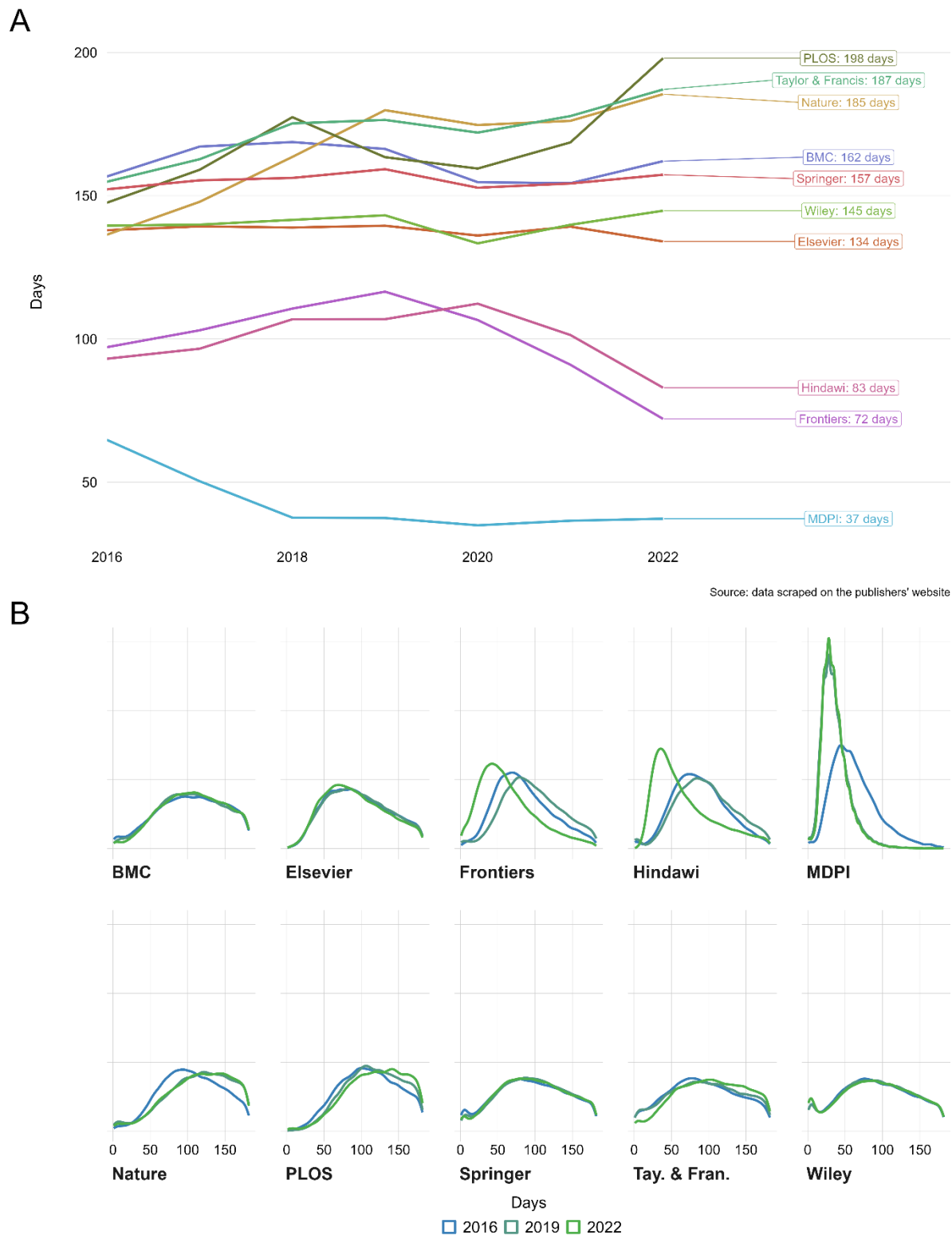
Figure 3: Article turnaround times. A) Evolution of mean turnaround times by publisher. Only articles with turnaround times between 1 day and 2 years were included. This filter was applied to remove data anomalies such as immediate acceptance or missing values that default to Jan 1st 1970 (the "Unix epoch"). B) Article turnaround time distribution curves from 2016-2022, focused on the first six months to better show trends. While most publishers have a right-skewed curve, the three publishers highlighted previously for increased special issue use have a left-skewed curve that only became more extreme over time. These data reflect only a fraction of total articles shown in Fig. 1, limited due to sampling methodology (see Table 1). Tay. & Fran. = Taylor & Francis.

Here we find that turnaround times differ by publisher, associated with use of the special issue publishing model. Variance in turnaround times also decreases for publishers alongside adoption of the special issue model. These results suggest that special issue articles are typically accepted more rapidly and in more homogenous timeframes than normal articles, which, to our knowledge, has never been formally described.

## Journal rejection rates and trends are publisher-specific

If a publisher lowers their article rejection rates, all else being equal, it will lead to more articles being published. Such changes to rejection rate might also mean more lower-quality articles are being published. Peer review is the principal method of quality control that defines science (31), and so publishing more articles with lower quality may add to strain and detract from the meaning and authority of the scientific process.

The relationship between rejection and quality is complex. High rejection rates do not necessarily reflect greater rigour: rigorous science can be rejected if the editors think that the findings lack the scope required for their journal. Conversely, low rejection rates may reflect a willingness to publish rigorous science independent of scope. The publisher also defines what "rejection rate" means in-house, creating caveats for comparing raw numbers across publishers.

Rejection rate data are rarely made public, and only a minority of publishers provide these data routinely or shared rejection rates upon request. Using the rejection rate data we could collect, we estimated rejection rates per publisher and asked if they: 1) change with growth in articles, 2) correlate with journal size, 3) predict article turnaround times, 4) correlate with journal impact, 5) depend on the publisher, or 6) predict a journal's proportion of species issue articles.

We found no clear trend between the evolution of rejection rates and publisher growth (Fig. 4A). Focussing on younger journals (≤10 years, ensuring fair comparisons) we found no relationship between journal size and reported or calculated 2022 rejection rates (Fig. 4B). Turnaround times are also not a strong predictor of rejection rates (Fig. 4supp1A). Finally, citations per document (similar to Clarivate IF) did not correlate with rejection rates (Fig. 4supp1B), indicating citations are not a strong predictor. Ultimately, the factor that best predicted rejection rates was the publisher itself: although both Frontiers and MDPI have similar growth in special issue articles (Fig. 2), they show opposite trends in rejection rates over time, and MDPI uniquely showed decreasing rates compared publishers (Fig. 4A). Raw rejection rates for MDPI in 2022 were also lower than other publishers. Curiously, Hindawi and MDPI journals with more special issue articles also had lower rejection rates ($P = 5.5e-8$ and $P = .01$ respectively, Fig. 4supp2), which we could not assess for other publishers.

In summary we found no general associations across publishers between rejection rate and most other metrics we investigated. Over time or among journals of similar age, rejection rate patterns were largely publisher-specific. We did, however, recover a trend that within publishers, rejection rates decline with increased use of special issue publishing.
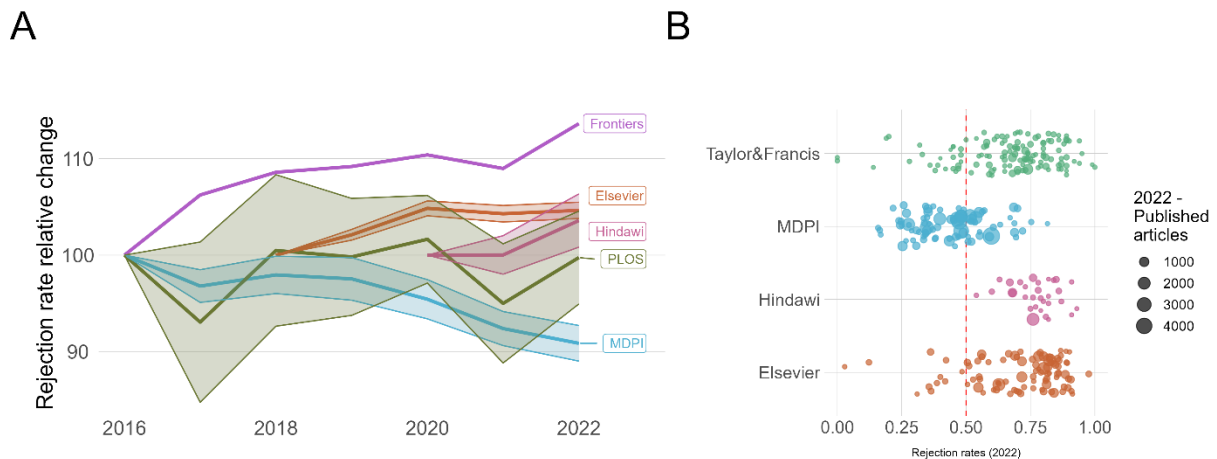
Figure 4: rejection rates are defined most specifically by publisher. A) Rejection rates differ markedly across publisher, including trends of increase, decrease, or no change. We estimated publisher rejection rates from varying available data, so we normalised these data by setting the first year on record as "100." Within publisher, we assume underlying data and definitions of 'rejection' are consistent from year to year, allowing comparisons among trends themselves. Frontiers data are the aggregate of all Frontiers journals, preventing the plotting of 95% confidence intervals. B) 2022 rejection rates among young journals (<10 years old) differ by publisher, but not journal size (B) or other metrics (Fig. 4supp1).

Disproportionately inflated Impact Factor affects select publishers

Among the most important metrics of researcher impact and publisher reputation are citations. For journals, the Clarivate 2-year IF reflects the mean citations per article in the two preceding years. Here we found that IF has increased across publishers in recent years (Fig. 5supp1,2). Explaining part of this IF inflation, we observed an exponential increase in total references per document between 2018-2021 (Fig. 5supp3, and see (30)). However, we previously noted that IF is used as a "badge of quality" by both researchers and publishers to earn prestige, and that IF can be abused by patterns of self-citation. We therefore asked if changes in journal citation behaviour may have contributed to recent inflation of the IF metric.

To enable systematic analysis, we used Cites/Doc from the Scimago database as a proxy of Clarivate IF (Cites/Doc vs. IF: R^2 = 0.77, Fig. 5supp4A). We then compared Cites/Doc to the network-based metric "Scimago Journal Rank" (SJR). Precise details of these metrics are discussed in the supplementary methods (Supplementary Table 1 and see (18)). A key difference between SJR and Cites/Doc is that SJR has a maximum amount of 'prestige' that can be earned from a single source. As such, within-journal self-citations or citation cartel-like behaviour is rewarded in Cites/Doc and IF, but not SJR. We define the ratio of Cites/Doc to SJR (or IF to SJR) as "impact inflation."

Impact inflation differs dramatically across publishers (Fig. 5A), and has also increased across publishers over the last few years (Fig. 5supp5A). In 2022, impact inflation in MDPI and Hindawi were significantly higher than all other publishers (P^{adj} < .05). Interestingly, Frontiers had low impact inflation comparable to other publishers, despite growth patterns similar to MDPI and Hindawi.

The reason behind MDPI's anomalous impact inflation appears to be straightforward: MDPI journals nearly universally spiked in rates of within-journal self-citation during the study period
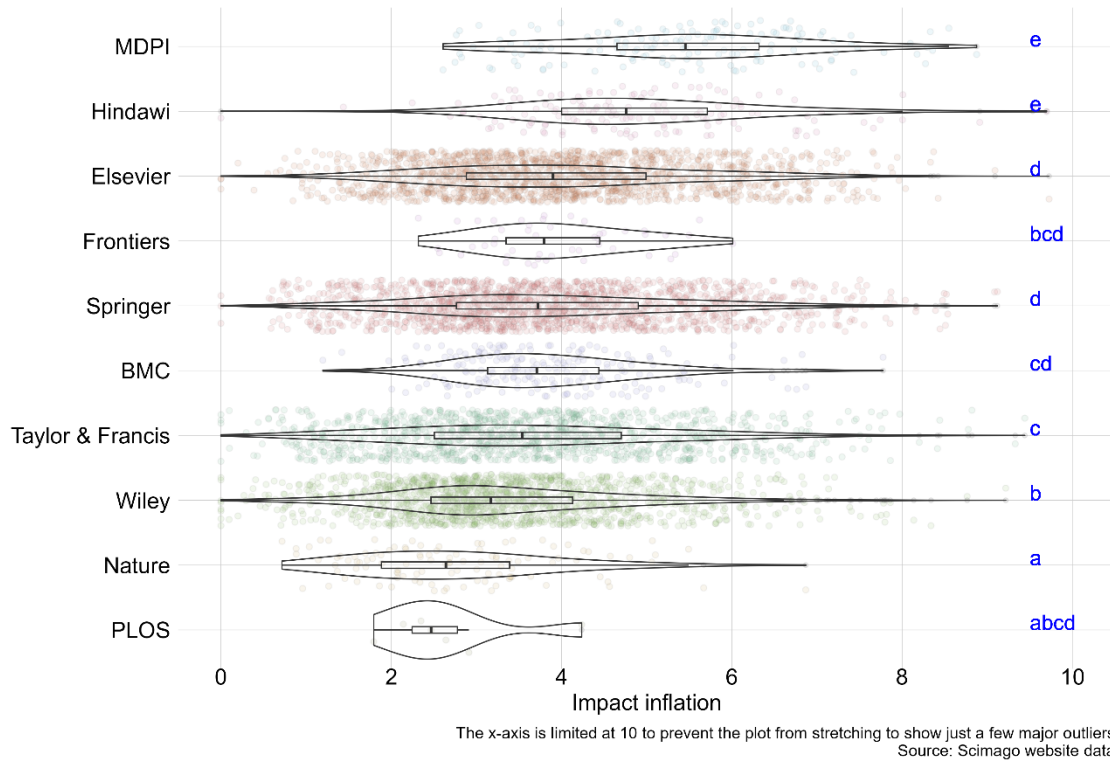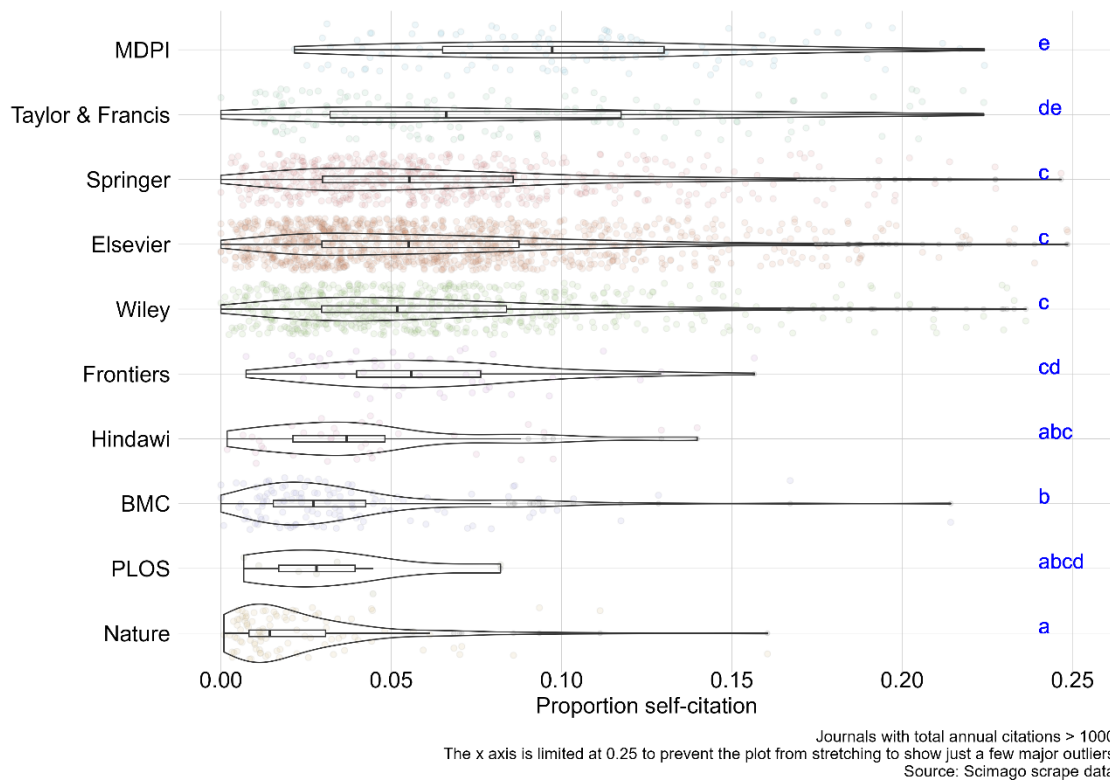
Figure 5: Changing behaviour of citation metrics revealed by Impact Inflation. Statistical letter groups reflect differences in one-way ANOVA with Tukey HSD. A) MDPI and Hindawi have significantly higher impact inflation compared to all other publishers. Comparisons using samples of Clarivate IFs are shown in Fig. 5supp4. B) MDPI journals have the highest rate of within-journal self-citation among compared publishers, including in previous years (Fig. 5supp5,6). Here we specifically analyse journals receiving at least 1000 citations per year to avoid comparing young or niche journals to larger ones expected to have diverse citation profiles.

Acronyms

Impact Factor (IF), Scimago Journal Rank (SJR)

(Fig. 5supp5B), with significant differences in self-citation rates compared to other publishers (Fig. 5B, P^{adj} < .05 and MDPI vs. Taylor & Francis, P^{adj} = .13), including comparisons in previous years (5supp6, $P^{adj\,2021}$ < .05, and MDPI vs. Taylor & Francis $P^{adj\,2021}$ =3e-7). Indeed, beyond within-journal self-citations, in an analysis from 2021, MDPI journals received ~29% of their citations from other MDPI journals (31), which would be rewarded per citation for IF but not SJR. Notably, Hindawi had self-citation rates more comparable to other publishers (Fig. 5B, Fig. 5supp6), despite high impact inflation. In this regard, while Hindawi journals may not directly cite themselves as often, they may receive many citations from a small network of journals, including many citations from MDPI journals (example in Fig. 5supp7).

In summary, we provide a novel metric, "impact inflation," that uses publicly-available data to assess journal citation behaviours. Impact inflation describes how proportionate a journal's total citations are compared to a network-adjusted approach. In the case of MDPI, there was also a high prevalence of within-journal self-citation, consistent with reports by Oviedo-Garcia (32) and MDPI itself (31). However high impact inflation and self-citation is not strictly correlated with other metrics we have investigated.

Discussion

Here we have characterised the strain on scientific publishing, as measured by the exponential rise of indexed articles and the resulting inability of scientists to keep up with them. The collective addition of nearly one million articles per year over the last 6 years alone costs the research community immensely, both in writing and reviewing time and in fees and article processing charges. Further, given our strict focus on indexed articles, not total articles, our data likely underestimate the true extent of the strain – *the problem is even worse than we describe.*

The strain we characterise is a complicated problem, generated by the interplay of different actors in the publishing market. Funders want to get the best return on their investment, while researchers want to prove they are a good investment. The rise in scientific article output is only possible with the participation of researchers who act as authors, reviewers and editors. Researchers do this because of the "publish or perish" imperative (19), which rewards individual researchers who publish as much as possible, forsaking quality for quantity. On the other hand, publishers host and spur the system's growth in their drive to run a successful business. Publishers structure the market, control journal reputation, and as such are focal players – which has led to concerns regarding to what extent publisher behaviour is motivated by profit (28). Growth in published papers should be possible and could be welcome. However, in the business of science publishing, growth should never come at the cost of the scientific process.

Considering our metrics in combination (Table 2) also allows us to identify common trends and helps to characterise the role that different publishers play in generating the strain. Across publishers, article growth is the norm, with some groups contributing more than others. Impact factors and impact inflation have both increased universally, exposing the extent to which the publishing system itself has succumbed to *Goodhart's law*. Nonetheless, the vast majority of growth in total indexed articles has come from just a few publishing houses following two broad models.

Strain indicators at a glance: 2022 and evolution 2016-22

	2022					CHANGE 2016-22				
	TOTAL PAPERS	SHARE SPECIAL ISSUE	TURNAROUND TIME (DAYS)	REJECTION RATE	IMPACT INFLATION	NUMBER PAPERS	SHARE SPECIAL ISSUE	TURNAROUND TIME (DAYS)	REJECTION RATE	IMPACT INFLATION
Overall	2816k	38%	116	62%	3.3	+47%	+27pp	-23	-1pp	+1.1
Elsevier	498k	--	134	71%	4.0	+41%	--	-4	+5pp*	+1.5
MDPI	264k	88%	37	40%	5.4	+1080%	+14pp	-28	-8pp	+2.2
Springer	250k	3%	157	--	3.9	+52%	-1pp	+5	--	+1.5
Wiley	231k	5%	145	--	3.3	+36%	-2pp	+5	--	+1.2
Frontiers	114k	69%	72	48%	4.0	+675%	+20pp	-25	+14pp	+1.8
Taylor & Francis	105k	--	--	--	3.7	+59%	--	--	--	+1.5
Nature	57k	11%	185	--	2.8	+32%	+6pp	+49	--	+1
BMC	44k	10%	162	--	3.9	+73%	+1pp	+5	--	+1.5
Hindawi	39k	62%	83	74%	5.0	+139%	+36pp	-10	+3pp°	+1.9
PLOS	19k	1%	198	59%	2.6	-23%	-3pp	+50	-4pp	+1.1

For older publishing houses (e.g. Elsevier, Springer), growth was not driven by major growth across all journals, but by the synergy of mild growth in both total journals and articles per journal in tandem. Another strategy used only by certain for-profit, gold open access publishers, consisted in an increased use of special issue articles as a primary means of publishing. This trend was coupled with uniquely reduced turnaround times, and in specific cases, high impact inflation and reduced rejection rates. Despite their stark differences, the amount of strain generated through these two strategies is comparable.

The rich context provided by our metrics also provides unique insights. Ours is the first study, of which we are aware, to document that special issue articles are systematically handled differently from normal submissions: special issues have lower rejection rates, and also both lower and seemingly more homogeneous turnaround times. We also highlight the unique view one gets by considering different forms of citation metrics, and develop impact inflation (IF/SJR) as a litmus test for journal reputation, informing not on journal impact itself, but rather whether a journal's impact is proportional to its expected rank absent the contribution of e.g. citation cartels.

Throughout our study MDPI was an outlier in every metric – often by wide margins. MDPI had the largest growth of indexed articles (+1080%) and proportion of special issue articles (88%), shortest turnaround times (37 days), decreasing rejection rates (-8 percentage points), highest impact inflation (5.4), and the highest within-journal mean self-citation rate (9.5%). Ours is not the first study analysing MDPI (13, 32, 33), but our broader context highlights the uniqueness of their profile and of their contribution to the strain.

Some metrics appear to be principally driven by publisher's policies: rejection rates and turnaround time means and variances are largely independent from any other metric we assayed. This raises questions about the balance between publisher's oversight and scientific editorial independence. This balance is essential to maintain scientific integrity and authority: oversight should be sufficient to ensure rigorous standards, but not so invasive as to override the independence of editors. *Understanding how editorial independence is maintained in*

current publishing environments, though beyond the scope of this paper, is key to maintaining scientific integrity and authority.

Given the importance of scientific publishing, it is unfortunate that the basic data needed to inform an evidence-based discussion are so hard to collect. This discussion on academic publishing would be easier if the metrics we collected were more readily available – we had to web scrape to obtain many pieces of basic information. The availability of our metrics could be encouraged by groups such as the Committee on Publication Ethics (34), which publishes guides on principles of transparency. We would recommend transparency for: proportion of articles published through special issues (or other collection headings), article turnaround times, and rejection rates. Rejection rates in particular would benefit from an authority providing a standardised reporting protocol, which would greatly boost the ability to draw meaningful information from them. While not a metric we analysed, it also seems prudent for publishers to be transparent about revenue and operating costs, given much of the funding that supports the science publishing system comes from taxpayer-funded or non-profit entities. Referees such as Clarivate should also be more transparent; their decisions can have a significant impact on the quality of publisher badges (see Table 1supp1 and (35)), and yet the reasoning behind these decisions is opaque.

Greater transparency will allow us to document the strain on scientific publishing more effectively. However, it will not answer the fundamental question: how should this strain be addressed? Addressing strain could take the form of grassroots efforts (e.g. researcher boycotts) or authority actions (e.g. funder or committee directives, index delistings). Researchers, though, are a disparate group and collective action is hard across multiple disciplines, countries and institutions. In this regard, funders can change the publish or perish dynamics for researchers, thus limiting their drive to supply articles. We recommend funders to review the metrics we define here and adopt policies such as narrative CVs that highlight researchers' best work over total volume (36), which mitigate publish or perish pressures. Indeed, researchers agree that changes to research culture must be principally driven by funders (37), whose financial power could also help promote engagement with commendable publishing practices.

Our study shows that regulating behaviours cannot be done at the level of publishing model. Gold open access, for example, does not necessarily add to strain, as gold open access publishers like PLOS (not-for-profit) and BMC (for-profit) show relatively normal metrics across the board. Rather our findings suggest that addressing strain requires action be taken to address specific publishers and specific behaviours. For instance, collective action by the researcher community, or guidelines from funders or ethics commitees, could encourage fewer articles be published through special issues, which our study suggests are not held to the same standard as normal issues. Indeed, reducing special issue articles would already address the plurality of strain being added.

Here we have characterised the strain on scientific publishing. We hope this analysis helps advance the conversation among publishers, researchers, and funders to reduce this strain and work towards a sustainable publishing infrastructure.

Acronyms

Impact Factor (IF), Scimago Journal Rank (SJR)

Acknowledgements

Author contributions

Web scraping was performed by PGB and PC, and Scimago data curation by MAH. Global doctorate and global researcher data curation was done by MAH and DB. Data analysis in R was done by MAH, PGB, and PC. Conceptualisation was performed collectively by MAH, PGB, PC, and DB. The initial article draft was written by MAH. All authors contributed to writing and revising to produce the final manuscript.

Acronyms

Impact Factor (IF), Scimago Journal Rank (SJR)

References

1. C. W. Fox, A. Y. K. Albert, T. H. Vines, Recruitment of reviewers is becoming harder at some journals: a test of the influence of reviewer fatigue at six journals in ecology and evolution. *Res Integr Peer Rev* **2**, 3, s41073-017-0027–x (2017).

2. C. J. Peterson, C. Orticio, K. Nugent, The challenge of recruiting peer reviewers from one medical journal's perspective. *Proc (Bayl Univ Med Cent)* **35**, 394–396 (2022).

3. A. Severin, J. Chataway, Overburdening of peer reviewers: A multi-stakeholder perspective on causes and effects. *Learned Publishing* **34**, 537–546 (2021).

4. P. D. B. Parolo, *et al.*, Attention decay in science. *Journal of Informetrics* **9**, 734–745 (2015).

5. D. Maher, A. Aseffa, S. Kay, M. Tufet Bayona, External funding to strengthen capacity for research in low-income and middle-income countries: exigence, excellence and equity. *BMJ Glob Health* **5**, e002212 (2020).

6. G. Nakamura, B. E. Soares, V. D. Pillar, J. A. F. Diniz-Filho, L. Duarte, Three pathways to better recognize the expertise of Global South researchers. *npj biodivers* **2**, 17 (2023).

7. , U.S. scientific leaders need to address structural racism, report urges (2023) https:/doi.org/10.1126/science.adh1702 (May 6, 2023).

8. S.-N. C. Liu, S. E. V. Brown, I. E. Sabat, Patching the "leaky pipeline": Interventions for women of color faculty in STEM academia. *Archives of Scientific Psychology* **7**, 32–39 (2019).

9. E. Meijaard, M. Cardillo, E. M. Meijaard, H. P. Possingham, Geographic bias in citation rates of conservation research: Geographic Bias in Citation Rates. *Conservation Biology* **29**, 920–925 (2015).

10. A. Mlinarić, M. Horvat, V. Šupak Smolčić, Dealing with the positive publication bias: Why you should really publish your negative results. *Biochem Med (Zagreb)* **27**, 030201 (2017).

11. L. J. Hofseth, Getting rigorous with scientific rigor. *Carcinogenesis* **39**, 21–25 (2018).

12. D. Sarewitz, The pressure to publish pushes down quality. *Nature* **533**, 147–147 (2016).

13. A. Abalkina, Publication and collaboration anomalies in academic papers originating from a paper mill: Evidence from a Russia-based paper mill. *Learned Publishing*, leap.1574 (2023).

14. C. Candal-Pedreira, *et al.*, Retracted papers originating from paper mills: cross sectional study. *BMJ*, e071517 (2022).

15. H. Else, R. Van Noorden, The fight against fake-paper factories that churn out sham science. *Nature* **591**, 516–519 (2021).

16. A. Grudniewicz, *et al.*, Predatory journals: no definition, no defence. *Nature* **576**, 210–212 (2019).

17. E. Garfield, The History and Meaning of the Journal Impact Factor. *JAMA* **295**, 90 (2006).

18. V. P. Guerrero-Bote, F. Moya-Anegón, A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of Informetrics* **6**, 674–688 (2012).

19. D. R. Grimes, C. T. Bauch, J. P. A. Ioannidis, Modelling science trustworthiness under publish or perish pressure. *R. Soc. open sci.* **5**, 171511 (2018).

20. F. C. Fang, A. Casadevall, Research Funding: the Case for a Modified Lottery. *mBio* **7**, e00422-16 (2016).

21. D. Li, L. Agha, Big names or big ideas: Do peer-review panels select the best science proposals? *Science* **348**, 434–438 (2015).

22. W. Quan, B. Chen, F. Shu, Publish or impoverish: An investigation of the monetary reward system of science in China (1999-2016). *AJIM* **69**, 486–502 (2017).

23. M. Thelwall, *et al.*, In which fields do higher impact journals publish higher quality articles? *Scientometrics* **128**, 3915–3933 (2023).

24. M. Fire, C. Guestrin, Over-optimization of academic publishing metrics: observing Goodhart's Law in action. *GigaScience* **8**, giz053 (2019).

25. I. Fister, I. Fister, M. Perc, Toward the Discovery of Citation Cartels in Citation Networks. *Front. Phys.* **4** (2016).

26. E. A. Fong, R. Patnayakuni, A. W. Wilhite, Accommodating coercion: Authors, editors, and citations. *Research Policy* **52**, 104754 (2023).

27. (n.d.) Scimago, SCImago Journal & Country Rank [Portal] (2023).

28. J. P. A. Ioannidis, A. M. Pezzullo, S. Boccia, The Rapid Growth of Mega-Journals: Threats and Opportunities. *JAMA* **329**, 1253 (2023).

29. D. W. Grainger, Peer review as professional responsibility: A quality control system only as good as the participants. *Biomaterials* **28**, 5199–5203 (2007).

30. B. D. Neff, J. D. Olden, Not So Fast: Inflation in Impact Factors Contributes to Apparent Improvements in Journal Quality. *BioScience* **60**, 455–459 (2010).

31. MDPI, "Comment on: 'Journal citation reports and the definition of a predatory journal: The case of the Multidisciplinary Digital Publishing Institute (MDPI)' from Oviedo-García" (2021) (May 6, 2023).

32. M. Á. Oviedo-García, Journal citation reports and the definition of a predatory journal: The case of the Multidisciplinary Digital Publishing Institute (MDPI). *Research Evaluation* **30**, 405–419a (2021).

33. S. Copiello, On the skewness of journal self-citations and publisher self-citations: Cues for discussion from a case study. *Learned Publishing* **32**, 249–258 (2019).

34. E. Wager, The Committee on Publication Ethics (COPE): Objectives and achievements 1997–2012. *La Presse Médicale* **41**, 861–866 (2012).

Acronyms

Impact Factor (IF), Scimago Journal Rank (SJR)

35. MDPI, "Clarivate Discontinues IJERPH and JRFM Coverage in Web of Science" (2023) (May 21, 2023).

36. DORA, Changing the narrative: considering common principles for the use of narrative CVs in grant evaluation (2022) (July 31, 2023).

37. Wellcome Trust, "What Researchers Think About the Culture They Work In" (2020).

Acronyms

Impact Factor (IF), Scimago Journal Rank (SJR)

Supplementary Materials and Methods

Data collection

Global researcher statistics

Total PhD graduate numbers were obtained from the Organisation for Economic Co-operation and Development (OECD, https://stats.oecd.org) and filtered to remove graduates of "Arts and Humanities" to better focus on growth of graduates in Science, Technology, Engineering, and Mathematics (STEM). Other sources were consulted to complement OECD data with data for China and India (NSF, 2022; Zwetsloot et al., 2021). This choice was made to improve robustness by ensuring the inclusion of these two major populations did not affect data trends. However, these sources used independent parameters for assessment, and lacked data past 2019, and so while we could use estimates for China and India for 2020, we ultimately chose to show only the OECD PhD data in Fig. 1A. Figure 1supp1 considers the addition of these external data and includes projections to 2022 using quadratic regression given the plateauing trend.

We also compared total articles with the United Nations Educational, Scientific and Cultural Organization (UNESCO) data on researchers-per-million (full time equivalent) from the Feb 2023 release of the UNESCO Science, Technology, and Innovation dataset (http://data.uis.unesco.org, "9.5.2 Researchers per million inhabitants"). Figure 1supp1 considers these data, including projections to 2022, using a linear regression model given trends.

Only ~0.1% of journals in the overlap of the Scimago and Web of Science databases had their sole category listed as "Arts and Humanities," and so we ran analyses with or without those journals, which gave the same result. Strictly Arts and Humanities journals are not retained in our final datasets being analysed.

Publisher and journal-level data

Total articles published per year was obtained from Scimago (Scimago, 2023). Historical data (1999 to 2022) for total number of articles, total citations per document over 2 years, the Scimago Journal Rank (SJR) metric, and total references per document were obtained from Scopus via the Scimago web portal (https://www.scimagojr.com/journalrank.php). Scimago yearly data were downloaded with the "only WoS journals" filter applied to ensure the journals we include here were indexed by both Scopus and Web of Science (Clarivate). Within-journal self-citation rate was obtained from Scimago via web scraping.

Historical Impact Factor data (2012-2022) for a range of publishers (16,174 journals across BMC, Cambridge University Press, Elsevier, Emerald Publishing Ltd., Frontiers, Hindawi, Lippincott, MDPI, Springer, Nature, Oxford University Press, PLOS, Sage, Taylor & Francis,

and Wiley-Blackwell) was downloaded from Clarivate. Due to the download limit of 600 journals per publisher, these IFs represent only a subset of all IFs available.

Rejection rates were collected from publishers in a variety of ways: 1) obtained from online available publisher's reports (Frontiers: https://progressreport.frontiersin.org/peer-review)), 2) given upon request (PLOS, Taylor & Francis) and 3) web scraping of publicly-available data extracted from the journal or company websites (MDPI, Hindawi, Elsevier via https://journalinsights.elsevier.com/). Frontiers rejection rate data lack journal-level resolution, and are instead the aggregate from the whole publisher per year.

Article-level data

Several methods were used to obtain article submission and acceptance times (in order to calculate turnaround times), along with whether articles were part of a special issue (also called "Theme Issues", "Collections" or "Topics" depending on the publisher). PLOS, Hindawi and Wiley´s turnaround times were extracted directly from their corpus. The latter was shared with the authors by Wiley upon request, while PLOS' (https://plos.org/text-and-data-mining/) and Hindawi's (https://www.hindawi.com/hindawi-xml-corpus/) are available online. BMC, Frontiers, MDPI, Nature and Springer data were obtained via web scraping of individual articles and collecting data in "article information"-type sections. Taylor & Francis turnaround times were obtained via CrossRef ("CrossRef," 2023) by filtering all available ISSNs from Scimago. To obtain Elsevier turnaround times we first extracted all Elsevier related ISSNs from Scimago, queried these in CrossRef to obtain a list of DOIs, and then web scraped the data from those articles. We also collected information on whether Elsevier articles were part of special issues during our web-scraping. However, the resulting data were unusually spotty and incomplete: for instance, we had journals with only one article total with data on special issue status, which would falsely suggest that "100%" of articles in that journal were special issue articles. Ultimately we did not include Elsevier in our analysis of special issue articles.

Data analysis and rationale

Grouping of publishers per Scimago labels

Publisher labels in Scimago were aggregated according to key "brand" names such as "Elsevier" or "Springer"; e.g. Elsevier BV, Elsevier Ltd and similar were aggregated as Elsevier, or Springer GmbH & Co, Springer International Publishing AG as "Springer."

This does not entirely capture the nested publishing structures of certain "publishers" per Scimago labels. At the time of writing, Elsevier and Springer are both publishers who own >2500 journals according to self-descriptions. However, our dataset only assigns ~1600 journals to these publishers in 2022 (Fig. 1supp2). Reasons for this discrepancy between self-reported numbers and our aggregate numbers come from smaller, but independent, publisher groups operating under the infrastructure of these larger publishing houses. Two examples: **1)** Cell Press (> 50 journals) is owned by Elsevier, **2)** both BioMed Central (BMC, >200 journals) and Nature Portfolio (Nature, >100 journals) are owned by Springer. Ultimately, we

decided that publishing houses large enough to distinguish themselves with their own licensed names were managed and operated sufficiently independently from their parent corporations to be kept separate. Nevertheless, our dataset aggregates the majority of Elsevier and Springer journals under their namesakes, and so we feel the data we report are a representative sampling, even if we caution that interpretation of trends in "Elsevier", "Nature", or other publishers, should consider this caveat regarding nested publisher ownership.

Our choice of publishers to highlight in comparisons required careful judgements. Our goal of characterising strain meant that we had to focus on publishers that were sufficiently "large" as far as our strain metrics were concerned. We included publishers like Hindawi and Public Library of Science (PLOS) because they were uniquely 'large' in terms of certain business models. PLOS is the largest publisher in terms of articles per journal per year (Fig. 1C), while Hindawi is a major publisher in terms of publishing articles under the Special Issue model (Fig. 2) that is also of current public interest (Quaderi, 2023). We also retained BMC and Nature as independent entities in our study, as these publishers offer relevant comparisons among publishing models. BMC is a for-profit Open Access publisher that operates hundreds of journals, much like Hindawi, Frontiers, and MDPI. Nature is a hybrid model publisher that includes paywalled or Open Access articles, publishes more total articles than BMC (Fig. 1B), and was a distinct publishing group for which we could collect systematic data on Special Issue use and turnaround times (Fig. 2, Fig. 3). We were also able to collect a partial sampling of those data from Springer, but to merge the two would have caused Nature to contribute a strong plurality of trends in Springer data in Fig. 2 and Fig. 3, obscuring the trends of both this nested publishing house and of the remaining majority of Springer journals: indeed the proportion of Special Issues (Fig. 2), turnaround times (Fig. 3), Impact Inflation and self-citation (Fig. 4B) of Nature is significantly different from other Springer journals, sometimes by a wide margin.

In some cases, journal size was also a relevant factor for comparisons across publishers. As emphasised by the high number of articles per journal by PLOS, MDPI, and Frontiers (Fig. 1C), some publishers publish hundreds to thousands of articles per journal annually, while others publish far less. The age of journals was also tied to this article output, as newer journals publish fewer articles, but grow to publish thousands of articles annually in later years. We therefore considered journal size throughout (Fig. 1supp3), and in metrics like self-citations, which were censored for only journals receiving at least 1000 citations per year. These filters were applied to ensure comparisons across journals and publishers were being made fairly: for instance, small journals have fewer articles to self-cite to, and highly specific niche journals may have high rates of self-citation for sensible reasons. This was especially important for comparisons at the publisher level, as some publishers have increased their number of journals substantially in recent years (Fig. 1supp2), meaning a large fraction of their journals are relatively young and less characteristic of the publisher's trends according to their better-established journals.

Rejection rates

The analysis of rejection rates comes with important caveats: these data come from non-standardised data sources (each publisher decides how rejection rates are reported) and we use voluntarily-reported rather than systematic data (volunteer bias).

In most cases, publishers track the total submissions, rejections, and acceptances over a set period of time. This can sometimes be just a few months, or it can be the length of a whole year. We defined rejection rate as a function of accepted, rejected, and total submissions, depending on the data that were available for each publisher. However, this definition fails to account for the dynamic status of articles as they go through peer review. For instance, publishers may define "rejection" as any article sent back to the authors, even if the result was ultimately "accept." These differences can drastically affect the absolute value of rejection rates, as some publishers may count Schrödinger-esque submissions where the underlying article is tallied as both "rejected" and "accepted" with different timestamps.

We will also note that while Frontiers and MDPI provided their rejection rates publicly, we were forced to assemble their rejection rates manually. For Frontiers, we explicitly use *1 - (accepted articles / total submissions)*, however Frontiers reports an independent number they call "Rejected" articles that gives a lower number if used in the formula: *rejected articles / total submissions* (Frontiers data collected from https://progressreport.frontiersin.org/peer-review, accessed Sept. 4th, 2023). Meanwhile, MDPI rejection rate data were available via "Journal statistics" web page html code as "total articles accepted" and "total articles rejected." We therefore defined total submissions to MDPI as the sum of all accepted and rejected articles. On the other hand, Hindawi reported their rejection rates publicly on journal pages as "acceptance rate," although the underlying calculation method is not given.

Finally, rejection rates are intrinsically tied to editorial workload capacity and total submissions received. For some journals, total workload from submissions has trade-offs with what can be feasibly edited. For instance, eLife initially saw longer times to deciding on whether to desk reject an article or not during a trial that committed to publish all articles after peer review at the author's discretion (eLife, 2019). This change to longer processing times was presumably instinctive editor behaviour to avoid the ensuing workload of accepting all articles for peer review, and so the commitment of time associated. In other journals, relatively few articles per editor might be submitted, and so more articles could be considered for publication and retained for reassessment following revisions, perhaps visible as broader turnaround time distribution curves (Fig. 3B). Thus, we will stress that the absolute rejection rate itself is not a measure of quality or rigour, but rather reflects the balance between editorial capacity, journal scope and mission, and the total submissions received.

While we could not standardise the methodology used to calculate rejection rates across publishers, we make the assumption that publishers have at least maintained a consistent methodology internally across years. For this reason, while comparing raw rejection rates comes with many caveats, comparing the direction of change itself in rejection rates shown in Fig. 4A should be relatively robust to differences between groups.

Impact inflation

"Impact inflation" is a new synthetic metric we define, and so here we will take care to detail its characteristics, caveats, and assumptions in depth. Principally, this metric uses the ratio of the Clarivate Impact Factor (IF) to the Scimago Journal Rank (SJR).

One of the most commonly used metrics for judging journal impact is provided by Clarivate's annual Journal Citation Reports: the journal IF. Impact Factor is calculated as the mean total citations per article in articles recently published in a journal, most commonly the last two years. The formula for IF is as follows, where y represents the year of interest (Garfield, 2006):

$$IF_y = \frac{Total citations_y}{Total publications_{y-1} + Total publications_{y-2}}$$

The IF of a journal for 2022 is therefore:

$$IF_{2022} = \frac{Total citations_{2022}}{Total publications_{2021} + Total publications_{2020}}$$

The name "Impact Factor" refers to this calculation when done by Clarivate using their Web of Science database. However, the exact same calculation can be performed using other databases, including the Scopus database used by Scimago; indeed, these metrics are highly correlated (Fig5supp4A). Because of mass delistings by Clarivate in their 2023 Journal Citation Reports that affected many journals (Quaderi, 2023), which were not delisted in Scimago data, there is a decoupling of the Scimago Cites/Doc and Clarivate IF in 2022 data ($F = 6736$ on 1, 3544 df, adj-$R^2 = 0.72$) compared to previous years 2012-2021 ($F = 43680$ on 1, 13397 df, adj-$R^2 = 0.77$). The overall trends in Impact Inflation are robust to use of either Cites/Doc or IF in 2021 or 2022 (Fig. 5A, Fig. 5supp4). For continuity with discussion below, we will describe IF as a metric of journal "prestige," as IF is sometimes used as a proxy of journal reputation.

The Scimago Journal Rank (SJR) is a metric provided by Scimago that is calculated differently from Clarivate IF. The SJR metric is far more complex, and full details are better described in (Guerrero-Bote and Moya-Anegón, 2012). Here we will provide a summary of the key elements of the SJR that inform its relevance to Impact Inflation.

Supplementary Table 1: methodological differences between SJR and Impact Factor
(adapted from (Guerrero-Bote and Moya-Anegón, 2012))

	SJR	Impact Factor
Database	Scopus	Web of Science
Citation time frame	3 years	2 years
Self-citation contribution	Limited	Unlimited
Field-weighted citation	Weighted	Unweighted
Size normalisation	Citable document rate	Citable documents
Citation networks considered?	Yes	No

The SJR is principally calculated using a citation network approach (visualised in Fig. 5supp7). The reciprocal relationship of citations between journals is considered in the ultimate rank of SJR "prestige," including a higher value placed on citations between journals of the same

general field. The formula used by Scimago further limits the amount of prestige that one journal can transfer to itself or to another journal. This is explicitly described as a way to avoid *"problems similar to link farms with journals with either few recent references, or too specialized"* (Guerrero-Bote and Moya-Anegón, 2012); "link farms" are akin to so-called "citation cartels" described in (Abalkina, 2021; Fister et al., 2016). This is the most important difference between IF and SJR, as IF does not consider the source of where citations come from, while SJR does. As a result, SJR does not permit journals with egregious levels of self- or co-citation to inflate the ultimate SJR prestige value.

The ratio of IF/SJR can therefore reveal journals whose total citations (IF) come from disproportionately few citing journals. MDPI journals have a much lower SJR compared to their IF. The reason for this is exemplified from the ratios of citations in/out for the flagship MDPI journals International Journal of Molecular Sciences, International Journal of Environmental Research and Public Health, and Sustainability (see Fig. 5supp7). These three journals not only have high rates of within-journal self-citation (9.4%, 11.8%, 15.3% respectively in 2022), they and other MDPI journals further contribute the plurality of the total citations to each other (MDPI, 2021), and to other journals (e.g. Hindawi – BioMed Research International), which outside of the MDPI network are often not reciprocated (Fig. 5supp7).

Importantly, growth of articles per journal is not an intrinsic factor behind this disparity. Frontiers has seen a similar level of growth of its articles per journal as MDPI (Fig. 1C), enabled by using the special issues model (Fig. 2), but has far lower Impact Inflation scores (Fig. 5A). Frontiers also receives more diverse citations coming from a wider pool of journals, and only sparingly from other Frontiers journals (Fig. 5supp7). Importantly, we cannot comment on *why* these behaviours exist. What can be said is that the MDPI model of publishing seems to attract authors that cite within and across MDPI journals far more frequently than authors publishing with comparable for-profit Open Access publishers like Hindawi, Frontiers, or BioMed Central (BMC). Indeed, in a self-analysis published in 2021, MDPI's rates of within-publisher self-citation (~29%, ~500k articles) were highly elevated compared to other publishers of similar size (not an opinion shared by MDPI). Their rates were also higher than IEEE (~5%, ~800k articles), Wiley-Blackwell (~17%, ~1.2m articles), and Springer Nature (~24%, ~2.5m articles), lower only compared to Elsevier (~37%, ~3.1m articles) (MDPI, 2021).

Supplementary Figure 1: analysis of within-publisher self-citation rate performed by MDPI (MDPI, 2021) in response to Oviedo-Garcia (Oviedo-García, 2021). The original interpretation of this figure, as presented by MDPI, is: "It can be seen that MDPI is in-line with other publishers, and that its self-citation index is lower than that of many others; on the other hand, its self-citation index is higher than some others." Our data in Fig. 5B (2022) and Fig. 5supp6 (2021) suggest instead that established MDPI journals receiving >1000 citations per year have uniquely high rates of within journal self-citation, which are significantly different from other publishers. This filter for only journals receiving >1000 citations is key, as due to the growth of MDPI journals in recent years, not including this caveat can give the false impression that MDPI, overall, has comparable rates of within-journal self-citation due to the many recent journals with relatively few articles that cannot easily cite themselves (but *can* cite other MDPI journals).

The ratio of IF to SJR (or of the Scimago proxy Cites/Doc to SJR) therefore assesses how two different citation-based metrics compare. The first metric (IF) is source-agnostic, counts the raw volume of citations and documents, and outputs a prestige score. The second metric has safeguards built in that prevent citation cartel-like behaviour from inflating a journal's prestige, and so if a journal receives a large number of its citations from just a few journals, it will not receive an SJR score that is proportional to its IF.

Comment on the advertisement of IF as a metric of prestige

It is striking to note that most journals celebrate a year-over-year increase in IF, however our study shows that IF itself has become inflated, like a depreciating currency, by the huge growth in total articles and total citations within those articles (Fig. 5supp3, Fig. 5supp5). As a result,

unless IF is considered as a relative rank, the value of a given IF changes over time. Indeed, a publisher whose journals had an average Cites/Doc of "3.00" in 2017 was somewhat high within our publisher comparisons, however in 2022 a Cites/Doc of "3.00" is near the lowest average Cites/Doc across publishers (Fig. 5supp1). This rapid inflation, i.e. depreciation of IF-like metrics, does not affect the <u>relative</u> comparisons made in Clarivate Journal Citation Reports' IF rank or IF percentile. Publishers often report absolute IFs, however our analysis suggests the more accurate IF-based metric to report would be a relative rank, such as IF rank or percentile within a given category.

As our impact inflation metric is similarly proportional to IF itself, we would likewise recommend adaptations of impact inflation to compare relative ranks, such as quartiles. Unlike IF, the impact inflation metric already normalises by journal size and field by calculating SJR through citable document rate, rather than citable documents (Supplementary Table 1), making field-specific normalisation less important for comparisons of impact inflation across journals.

Data and code availability

At this time (Sept 2023) we were legally advised to withhold sharing our code and data files. Our work was done in accordance with UK government policy on text mining for non-commercial research (Gov.uk, 2021), and we anticipate being able to share much of code and data in the future. Code and data will be made available to peer reviewers to ensure a robust peer review process.

We used R version 4.3.1 (R Core Team 2023) and the following R packages: agricolae v. 1.3.6 (de Mendiburu 2023), emmeans v. 1.8.7 (Lenth 2023), ggtext v. 0.1.2 (Wilke and Wiernik 2022), gridExtra v. 2.3 (Auguie 2017), gt v. 0.9.0 (Iannone et al. 2023), gtExtras v. 0.4.5 (Mock 2022), here v. 1.0.1 (Müller 2020), hrbrthemes v. 0.8.0 (Rudis 2020), kableExtra v. 1.3.4 (Zhu 2021), magick v. 2.7.5 (Ooms 2023), MASS v. 7.3.60 (Venables and Ripley 2002), multcomp v. 1.4.25 (Hothorn, Bretz, and Westfall 2008), multcompView v. 0.1.9 (Graves, Piepho, and Sundar Dorai-Raj 2023), MuMIn v. 1.47.5 (Bartoń 2023), mvtnorm v. 1.2.2 (Genz and Bretz 2009), patchwork v. 1.1.2 (Pedersen 2022), scales v. 1.2.1 (Wickham and Seidel 2022), sjPlot v. 2.8.15 (Lüdecke 2023), survival v. 3.5.5 (Therneau T 2023), TH.data v. 1.1.2 (Hothorn 2023), tidyverse v. 2.0.0 (Wickham et al. 2019) and waffle v. 0.7.0 (Rudis and Gandy 2017).

Package citations

Auguie, Baptiste. 2017. gridExtra: Miscellaneous Functions for "Grid" Graphics. https://CRAN.R-project.org/package=gridExtra.

Bartoń, Kamil. 2023. MuMIn: Multi-Model Inference. https://CRAN.R-project.org/package=MuMIn.

de Mendiburu, Felipe. 2023. agricolae: Statistical Procedures for Agricultural Research. https://CRAN.R-project.org/package=agricolae.

Genz, Alan, and Frank Bretz. 2009. Computation of Multivariate Normal and t Probabilities. Lecture Notes in Statistics. Heidelberg: Springer-Verlag.
Graves, Spencer, Hans-Peter Piepho, and Luciano Selzer with help from Sundar Dorai-Raj. 2023. multcompView: Visualizations of Paired Comparisons. https://CRAN.R-project.org/package=multcompView.

Hothorn, Torsten. 2023. TH.data: TH's Data Archive. https://CRAN.R-project.org/package=TH.data.

Hothorn, Torsten, Frank Bretz, and Peter Westfall. 2008. "Simultaneous Inference in General Parametric Models." Biometrical Journal 50 (3): 346–63.

Iannone, Richard, Joe Cheng, Barret Schloerke, Ellis Hughes, Alexandra Lauer, and JooYoung Seo. 2023. gt: Easily Create Presentation-Ready Display Tables. https://CRAN.R-project.org/package=gt.

Lenth, Russell V. 2023. emmeans: Estimated Marginal Means, Aka Least-Squares Means. https://CRAN.R-project.org/package=emmeans.

Lüdecke, Daniel. 2023. sjPlot: Data Visualization for Statistics in Social Science. https://CRAN.R-project.org/package=sjPlot.

Mock, Thomas. 2022. gtExtras: Extending "gt" for Beautiful HTML Tables. https://CRAN.R-project.org/package=gtExtras.

Müller, Kirill. 2020. here: A Simpler Way to Find Your Files. https://CRAN.R-project.org/package=here.

Ooms, Jeroen. 2023. magick: Advanced Graphics and Image-Processing in r. https://CRAN.R-project.org/package=magick.

Pedersen, Thomas Lin. 2022. patchwork: The Composer of Plots. https://CRAN.R-project.org/package=patchwork.

R Core Team. 2023. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rudis, Bob. 2020. hrbrthemes: Additional Themes, Theme Components and Utilities for "ggplot2". https://CRAN.R-project.org/package=hrbrthemes.

Rudis, Bob, and Dave Gandy. 2017. waffle: Create Waffle Chart Visualizations in r. https://CRAN.R-project.org/package=waffle.

Therneau T (2023). A Package for Survival Analysis in R. https://CRAN.R-project.org/package=survival

Venables, W. N., and B. D. Ripley. 2002. Modern Applied Statistics with s. Fourth. New York: Springer. https://www.stats.ox.ac.uk/pub/MASS4/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." Journal of Open Source Software 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and Dana Seidel. 2022. scales: Scale Functions for Visualization. https://CRAN.R-project.org/package=scales.

Wilke, Claus O., and Brenton M. Wiernik. 2022. ggtext: Improved Text Rendering Support for "ggplot2". https://CRAN.R-project.org/package=ggtext.

Zhu, Hao. 2021. kableExtra: Construct Complex Table with "kable" and Pipe Syntax. https://CRAN.R-project.org/package=kableExtra.

Supplementary figures and tables
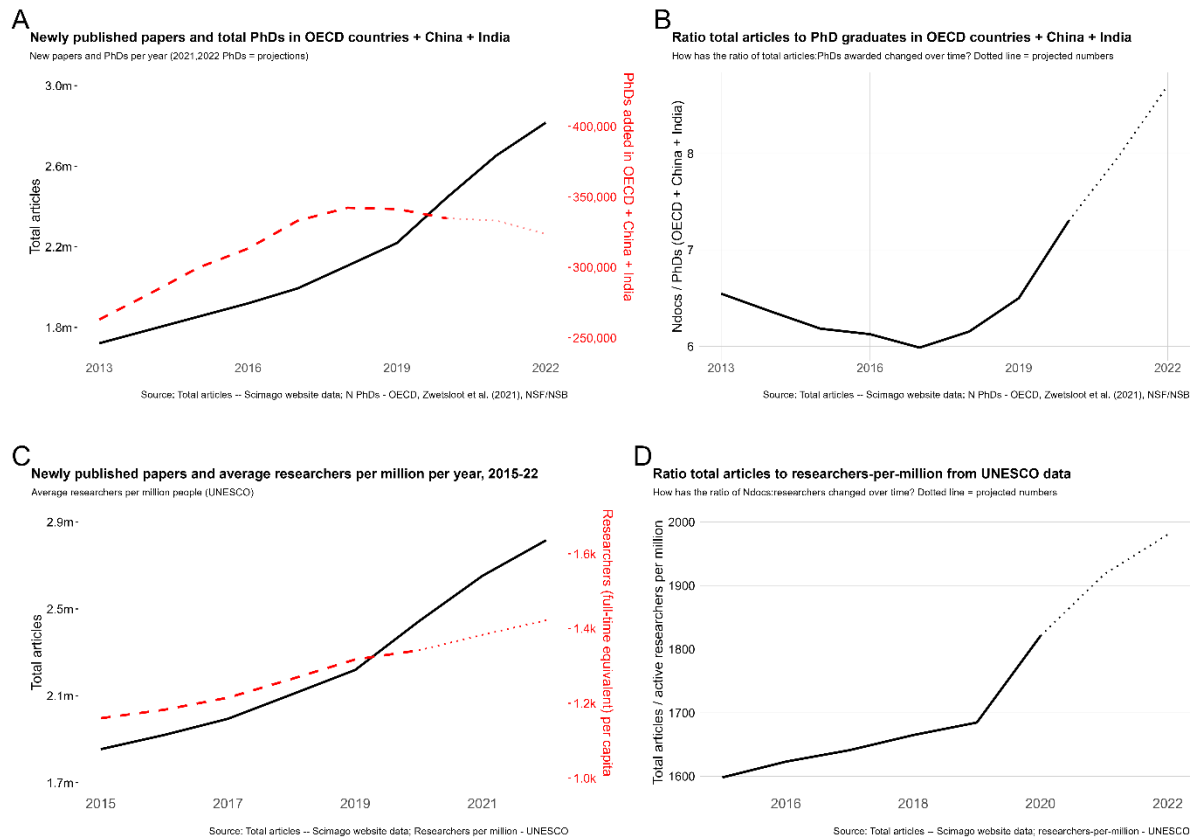
Fig1supp1: the growing disparity between total articles per year and active researchers is robust to use of alternate datasets. Dotted lines indicate estimated trends. A) OECD data complemented with total STEM PhD graduates from India and China (dashed red line) does not alter the pattern of an overall decline in recent years (Fig. 1A). B) The ratio of total articles to total PhD graduates has gone up substantially since 2019. C) UNESCO data instead using total active researchers (full-time equivalent) per million people shows a similar trend. Of note, this proxy for active researchers may include non-publishing scientists (private industry, governmental), that are not participating in the strain on scientific publishing in the same way academic scientists are. D) Nevertheless, using UNESCO data the ratio of total articles to total active researchers has gone up substantially since 2019.

Total journals by publisher
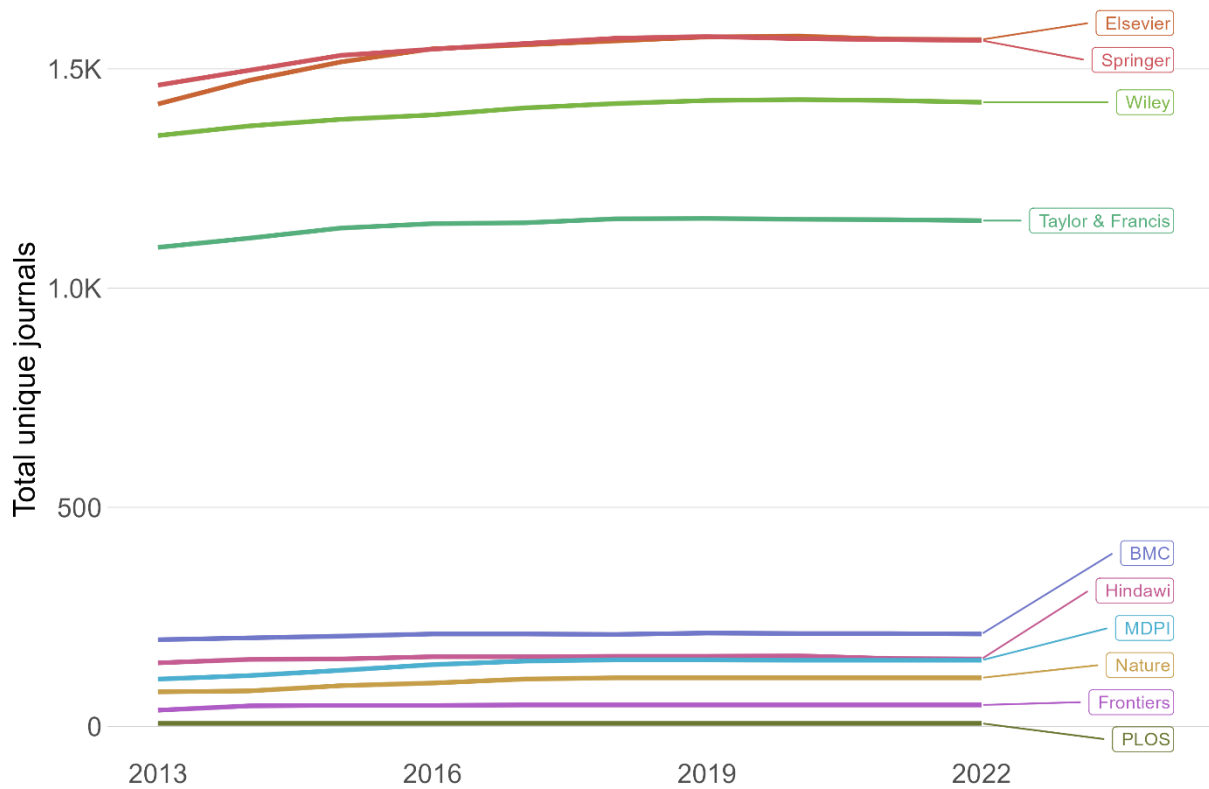
Source: Scimago website data

Fig1supp2: growth in total journals by publisher. Between 2013-2022, Elsevier, Springer, Taylor & Francis, MDPI, and Nature have added to their total journals noticeably. Note: we have only analysed journals indexed in both Scopus and Web of Science, and also collected journals under Publishers according to their licensed Publisher names. Subsidiary publishers under the umbrella of larger publishers are not included in larger publisher totals. For example, both BioMed Central (BMC) and Nature portfolio (Nature) are subsidiaries of Springer Nature (Springer), but host a large number of journals and license under a non-Springer name, and so are treated as separate entities in our study.
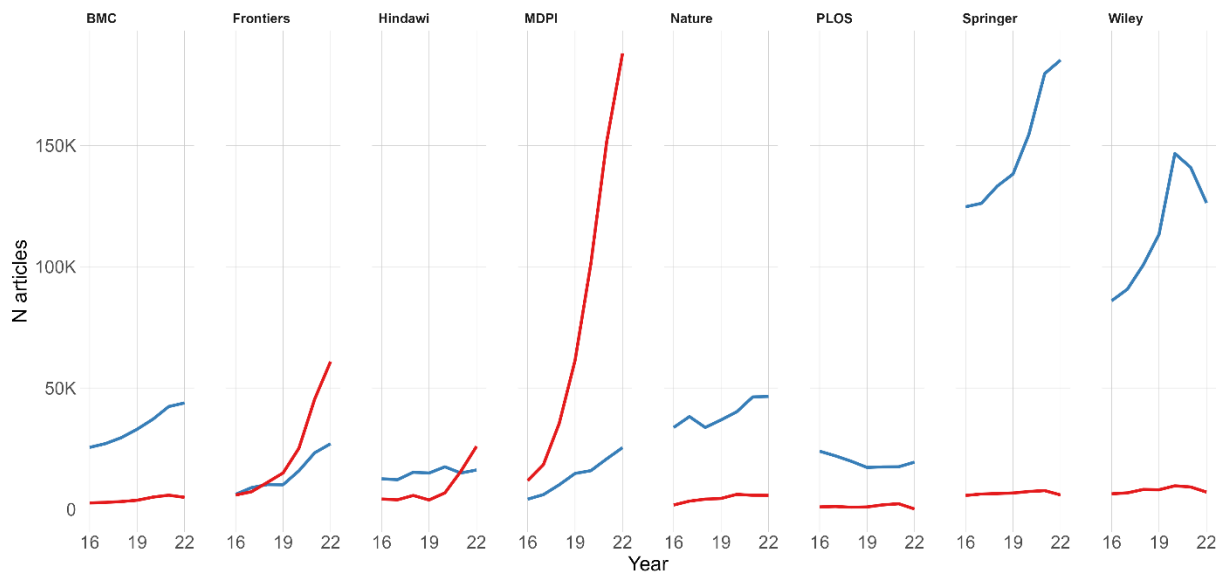
Number of journals by class of size, 2002-22
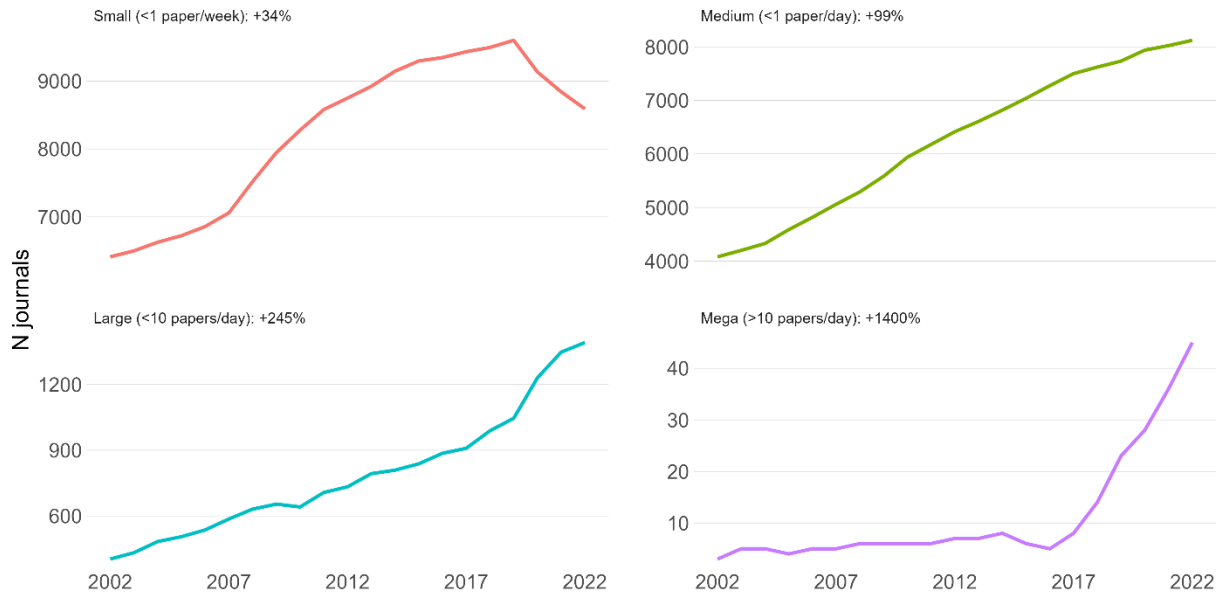
Fig1supp3: the rise of megajournals. We recover trends supporting the article by Ioannidis et al. *(28)* who described the "rise of megajournals." Specifically, we see a decline in the number of journals publishing <1 paper/week, but sharp increases in the number of journals publishing over a paper per day. Scientific publishing has therefore been concentrating more and more articles into fewer journals proportionally, which also coincides with a slight decline in the number of journals publishing only a few articles per year.

Number of papers published in regular vs special issues, 2016-22

Wiley decrease in 2022 likely due to limited coverage of Wiley papers in 2022

Source: data scraped from the publisher's website
Notes: Special issues are called Collections at PLOS and Topics at Frontiers. For MDPI Collections, Sections and Topics not shown.

Fig2supp1: proportion of articles published in regular vs. special issues. Underlying data are the same as in Fig. 2. Line plots are shown to better depict the year-by-year evolving proportion of special issue articles to regular articles. The decline in Wiley articles from 2020-2022 is an artefact of web scraping where total data availability declined in these years. As shown in Fig. 1B, Wiley overall article output increased slightly in recent years.

Evolution of the share of papers appearing in Special Issues, 2016 to 2022
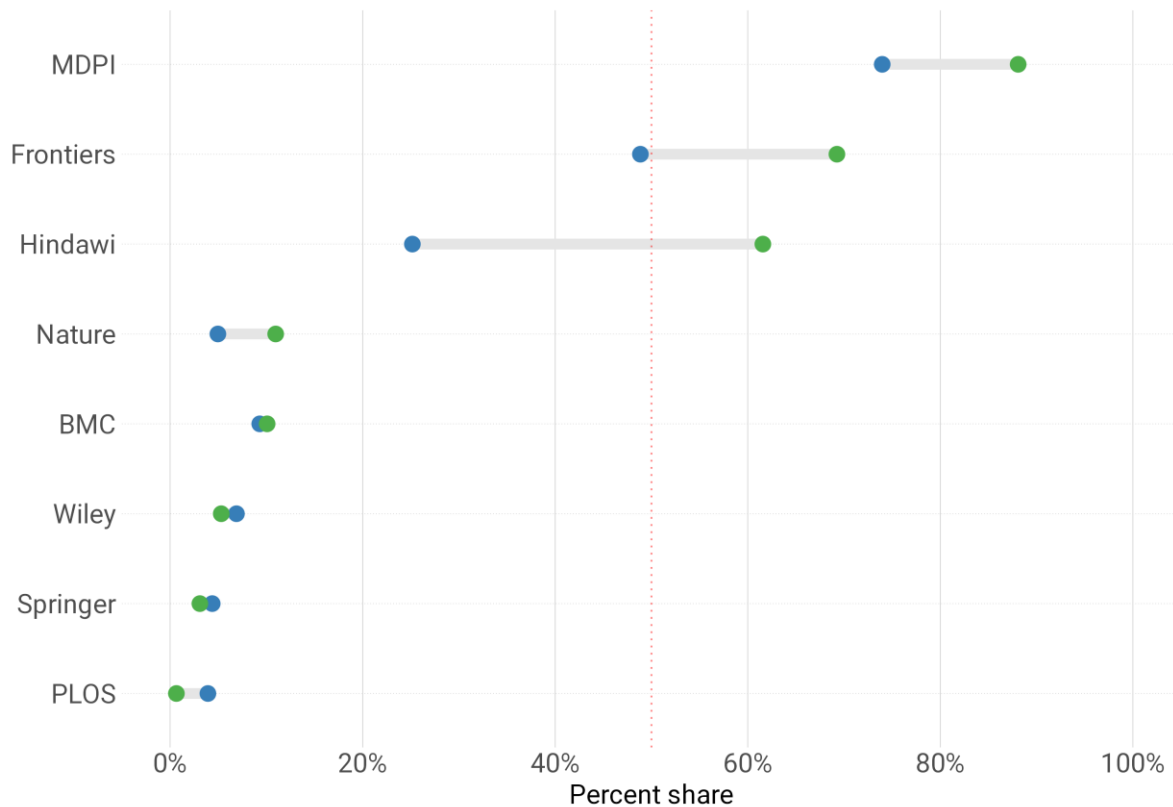
Source: data scraped from the publishers' website
Notes: Special issues are called Collections at PLOS and Topics at Frontiers. For MDPI Collections, Sections and Topics not shown.

Fig2supp2: change in special issue between 2016 and 2022. Certain groups publish the majority of their articles through special issues. Mean proportions of articles published through regular or special issues shown.

**Article heterogeneity in turnaround times by publisher, 2016-22**
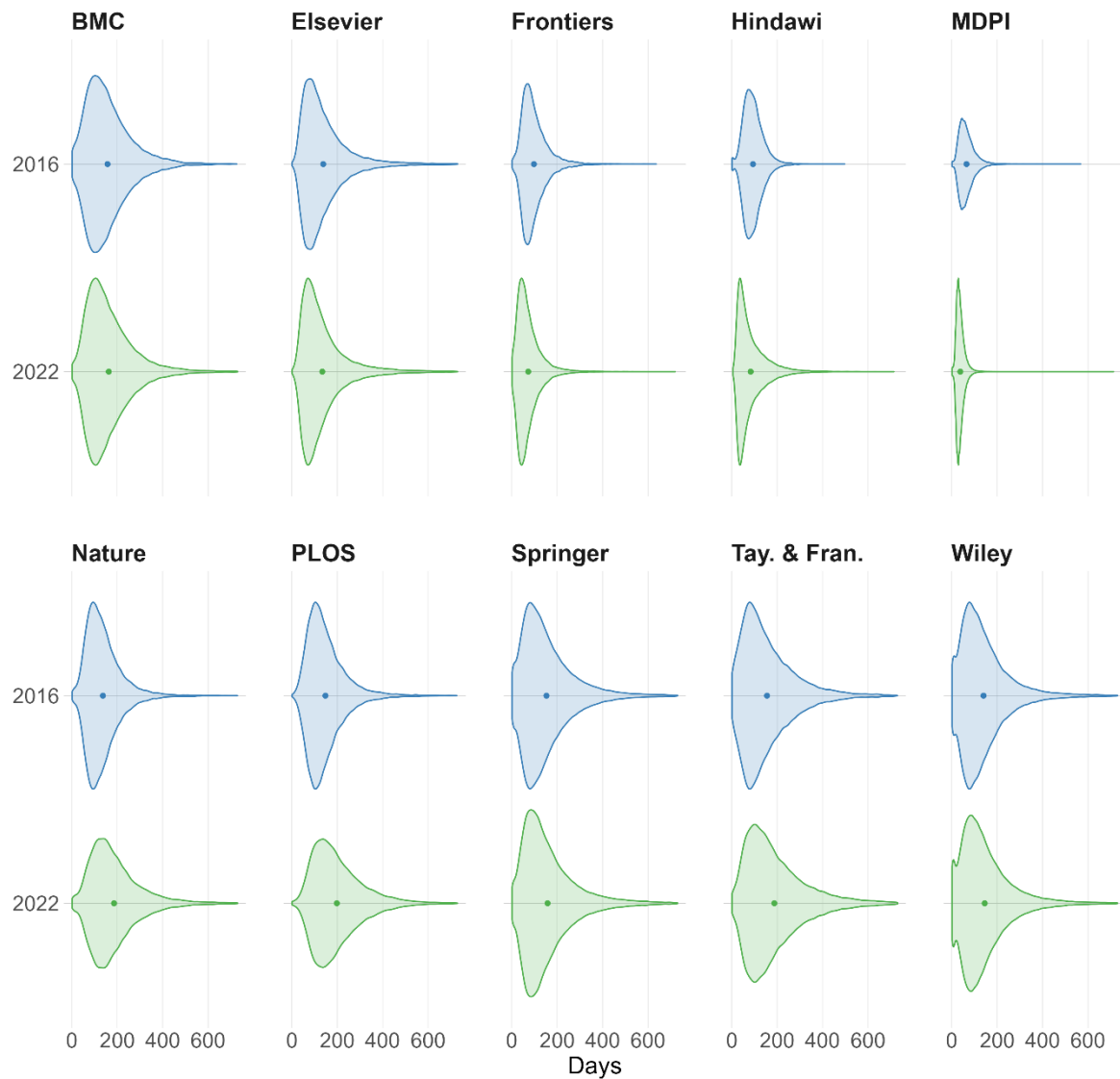
Source: data scraped from the publishers' websites

Fig3supp1: heterogeneity in journal mean turnaround times by publisher. Underlying data same as Fig. 3B. Here Violin plots provide an alternate depiction of the density of turnaround time distributions of all articles within their publishing house. "Tay. & Fran." = Taylor & Francis.
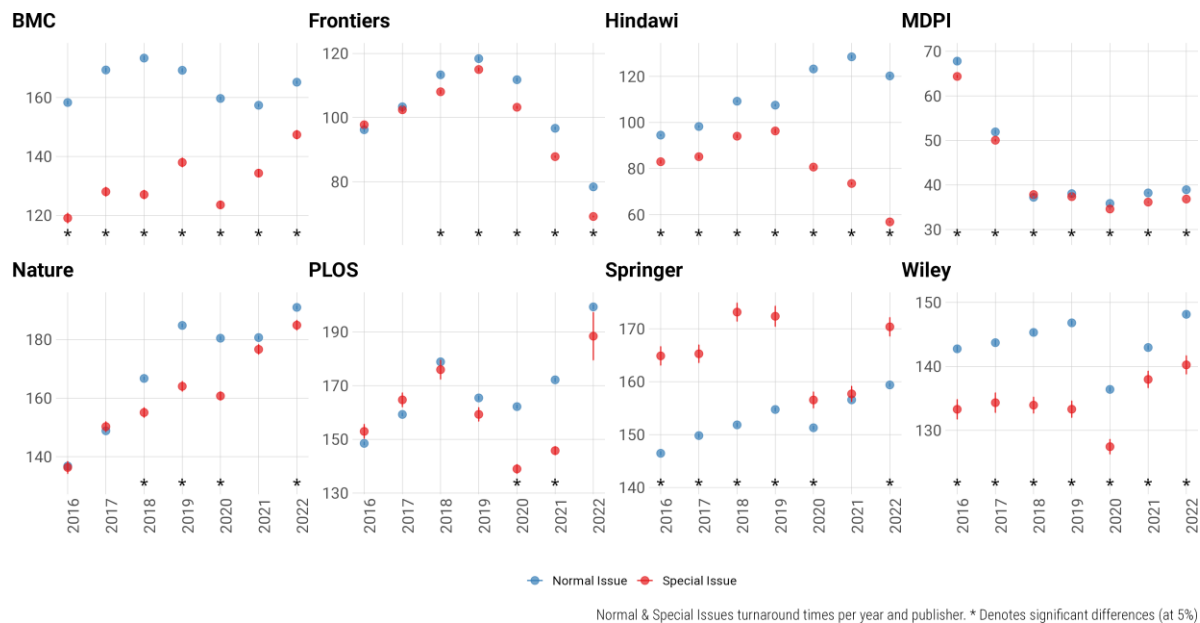
Fig3supp2: article turnaround times split by normal or special issue status. Across publishers, special issue articles have lower turnaround times, often by significant margins (for each year: Tukey HSD, p < .05 = *). The only exception to this trend is Springer, which had higher turnaround times for special issue articles. Of note, the way that special issues are organised can vary across journals and publishers, which could explain the differences in the extent of these trends by publisher. Error bars represent standard error.
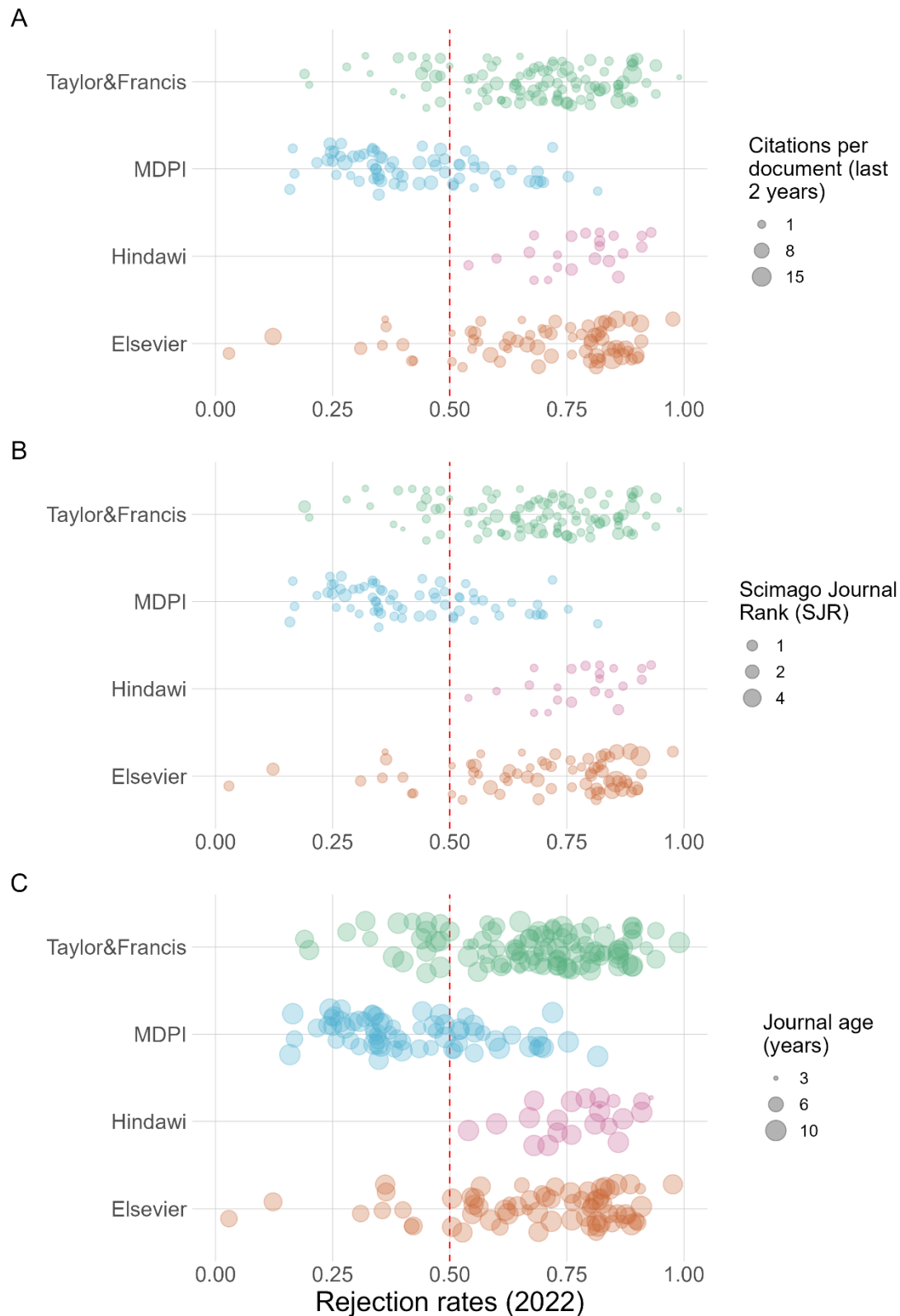
Fig4supp1: 2022 rejection rates by publisher, split across different parameters. Using a general linear model, we found no significant effect of total documents (Fig. 4B), citations per document (A), Scimago Journal Rank (B), or journal age (C) on a journal's 2022 rejection rates across publishers. We chose to investigate young journals (≤ 10 years) to avoid comparing long-established journals to new journals that might have different needs for growth.

Share of Special Issues and Rejection Rate at Hindawi and MDPI

92 MDPI journals with an IF as of January 2023, 72 Hindawi journals for which we have data
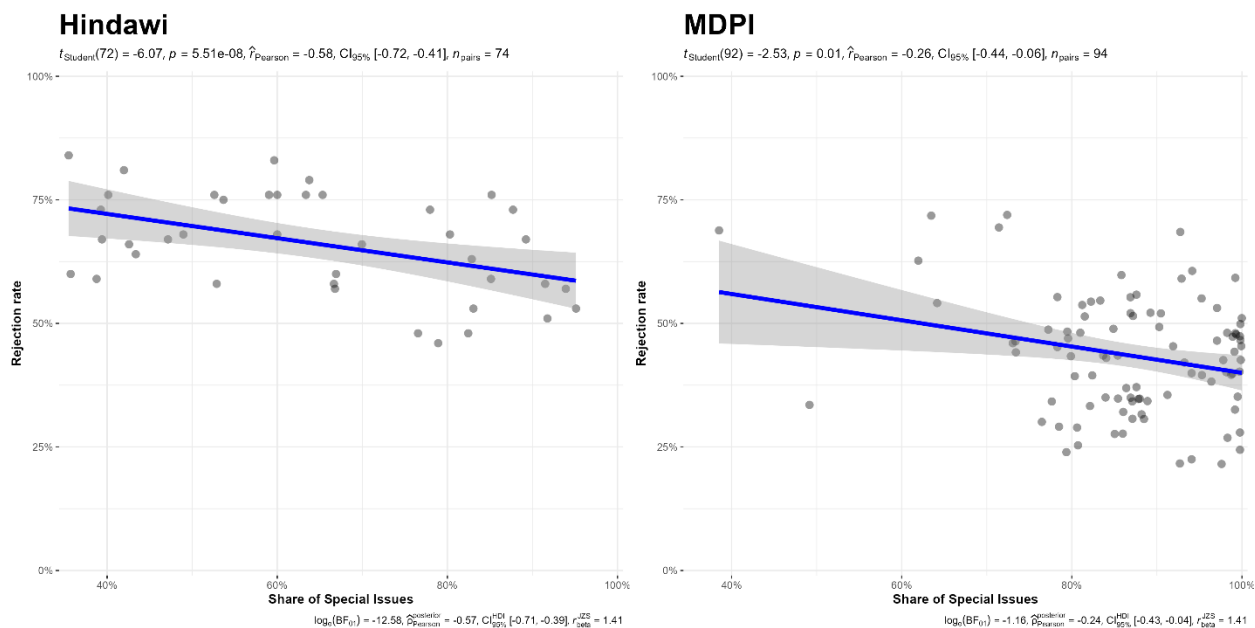
Fig4supp2: Rejection rates relative to proportions of special issue articles. For Hindawi and MDPI, two publishers that we could analyse, there was a significant correlation between 2022 journal rejection rates and their share of articles published through special issues.

Evolution of rejection rates by relative size of the journal at MDPI, 2016-22

Only journals existing in 2016

Number of articles per journal   5000   10000   15000
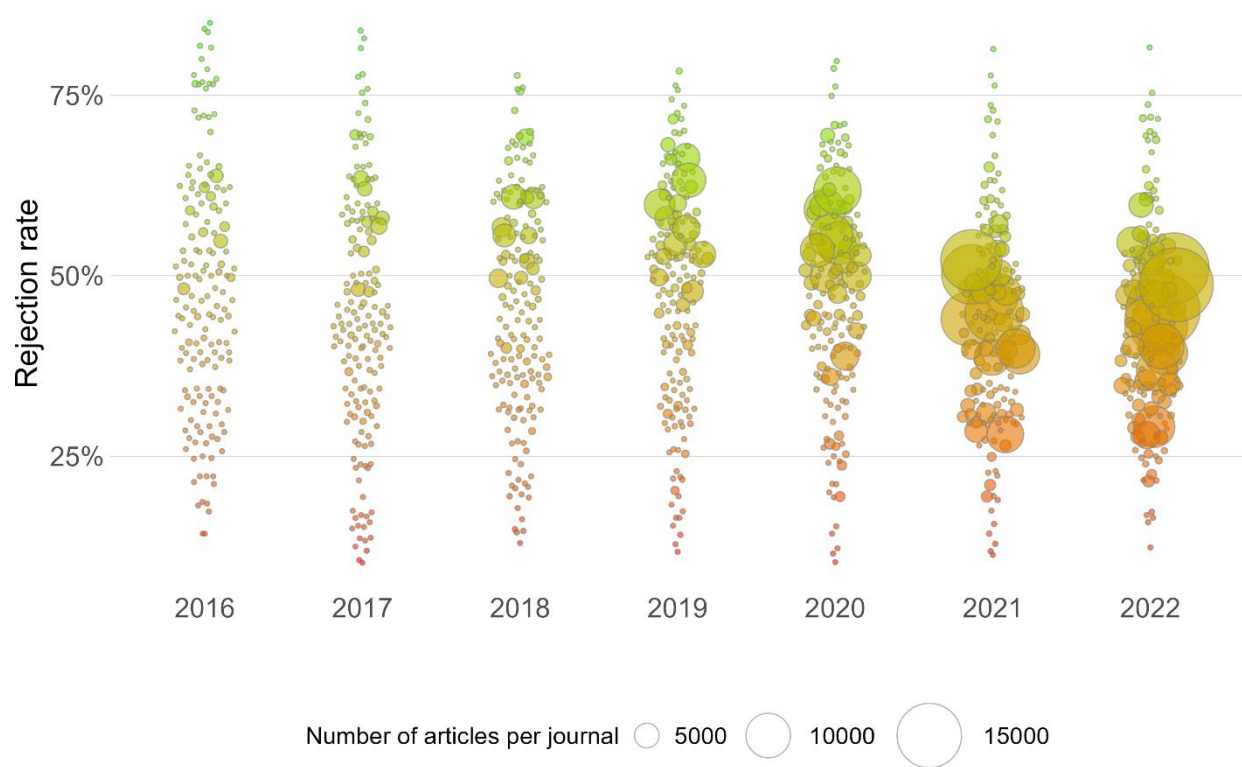
Source: data scraped on the publisher's website

Fig4supp3: the decline in MDPI rejection rates is present across journals of different size classes. A steady decline in rejection rates began between 2019-2020 (Fig. 4A) alongside growth in journal size (larger bubbles here).

A
Change in Scimago Cites/Doc (2years)

Cites/Doc (2 years) - proxy of IF

Nature
PLOS
Elsevier
BMC
Frontiers
MDPI
Wiley
Springer
Taylor & Francis
Hindawi

2013   2016   2019   2022

Source: Scimago website data

B
Change in SJR per publisher

Scimago Journal Rank (SJR)

Nature
PLOS
Elsevier
BMC
Wiley
Frontiers
Springer
MDPI
Taylor & Francis
Hindawi

2013   2016   2019   2022

Source: Scimago website data

Fig5supp1: raw Cites/Doc and SJR informing the Impact Inflation metric. A) Cites/Doc has been increasing year-over-year across publishers, with a notable uptick beginning after 2019. Here we describe the recent inflation of journal IF (with Cites/Doc as our proxy), suggesting the relative value of a given absolute IF number (e.g. "IF = 3") has decreased more rapidly than in years prior to 2019. B) The SJR has remained relatively constant in recent years, as expected since this metric is normalised for journal size and rate of citable documents generated, rather than raw total documents (see *(18)*). This suggests that the year-over-year increase in Impact Inflation (Fig. 5supp2) we've observed is due to increased total citations by increasing total articles, but those citations are not weighted as "prestigious" in a network-adjusted metric compared to pre-2019 years.
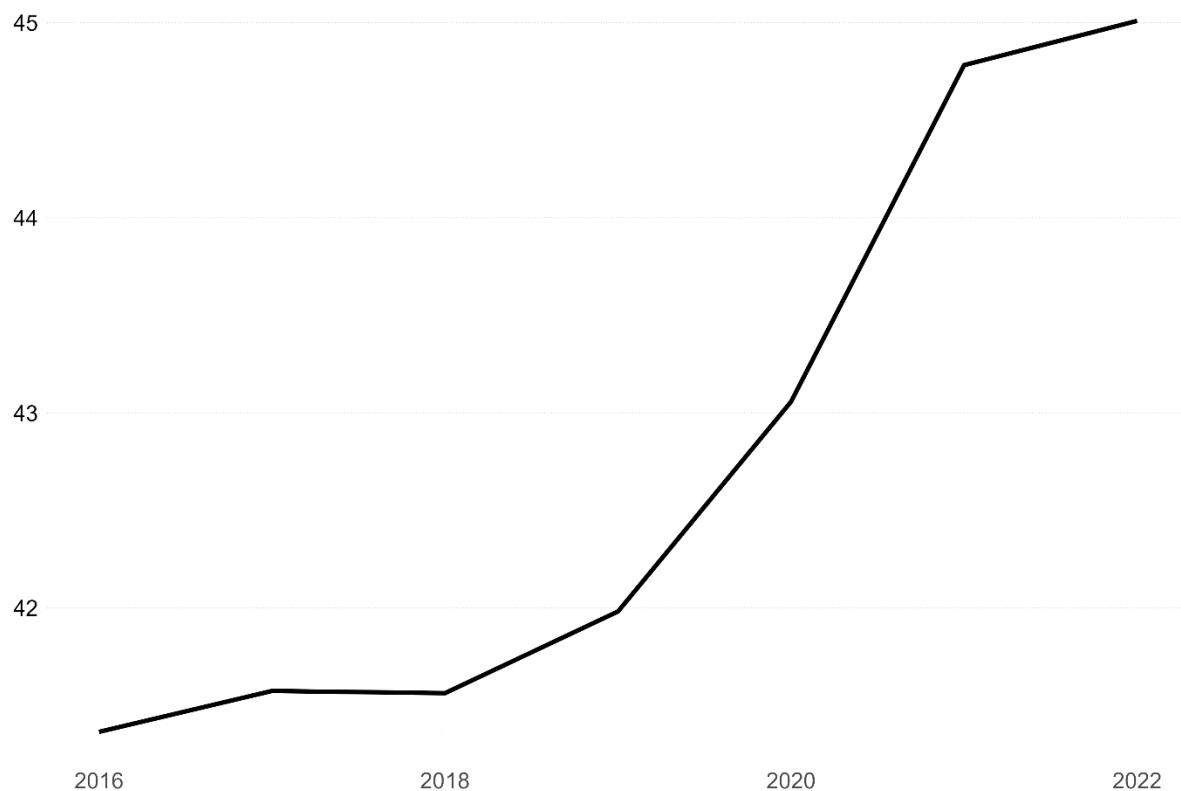
# Impact inflation by journal size 1999-2022

Cites/Doc (2 years) over SJR
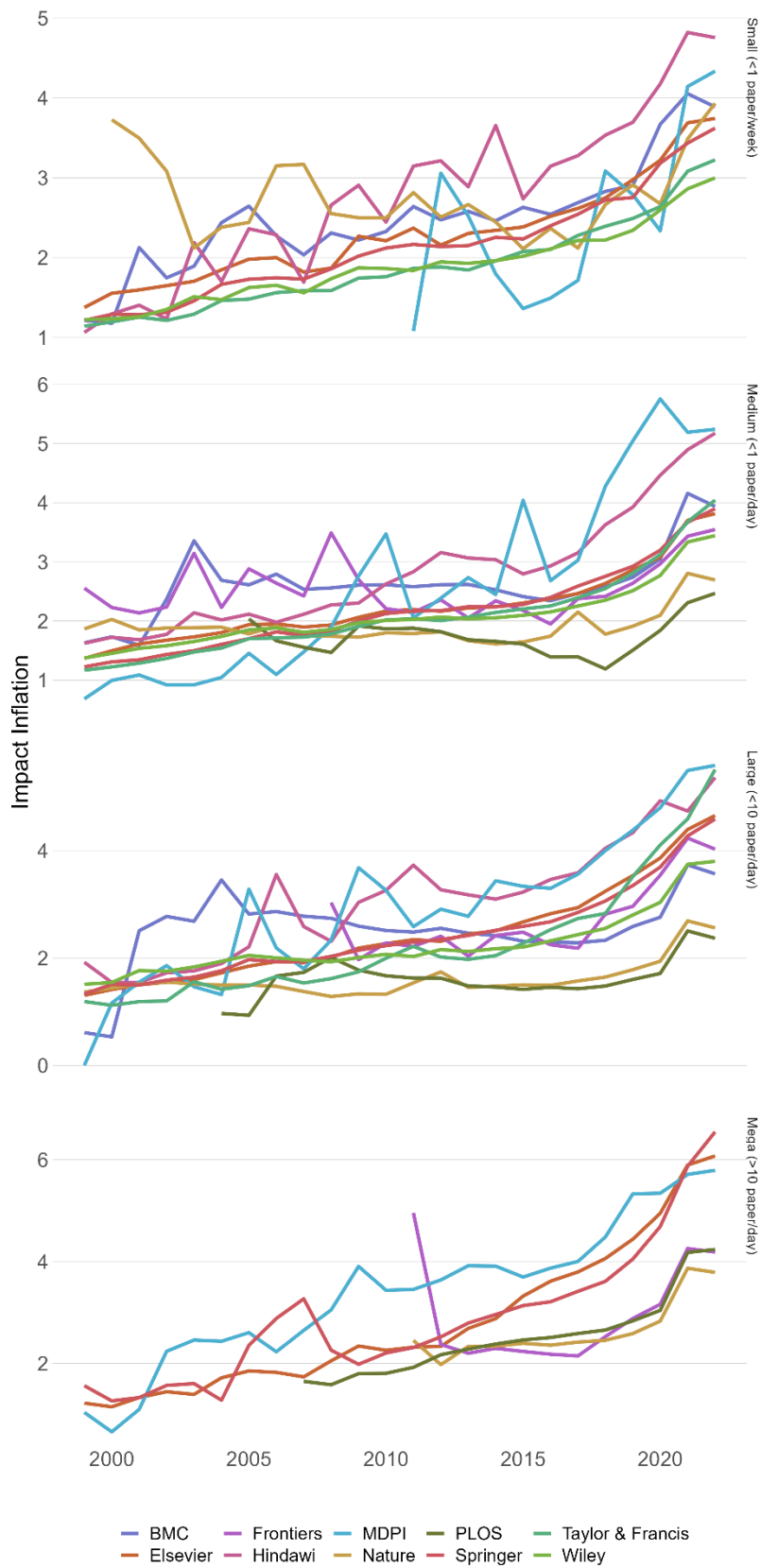
Fig5supp2: there has been a universal increase in Impact Inflation independent of journal size across all publishers. Also see Fig. 5supp5A.
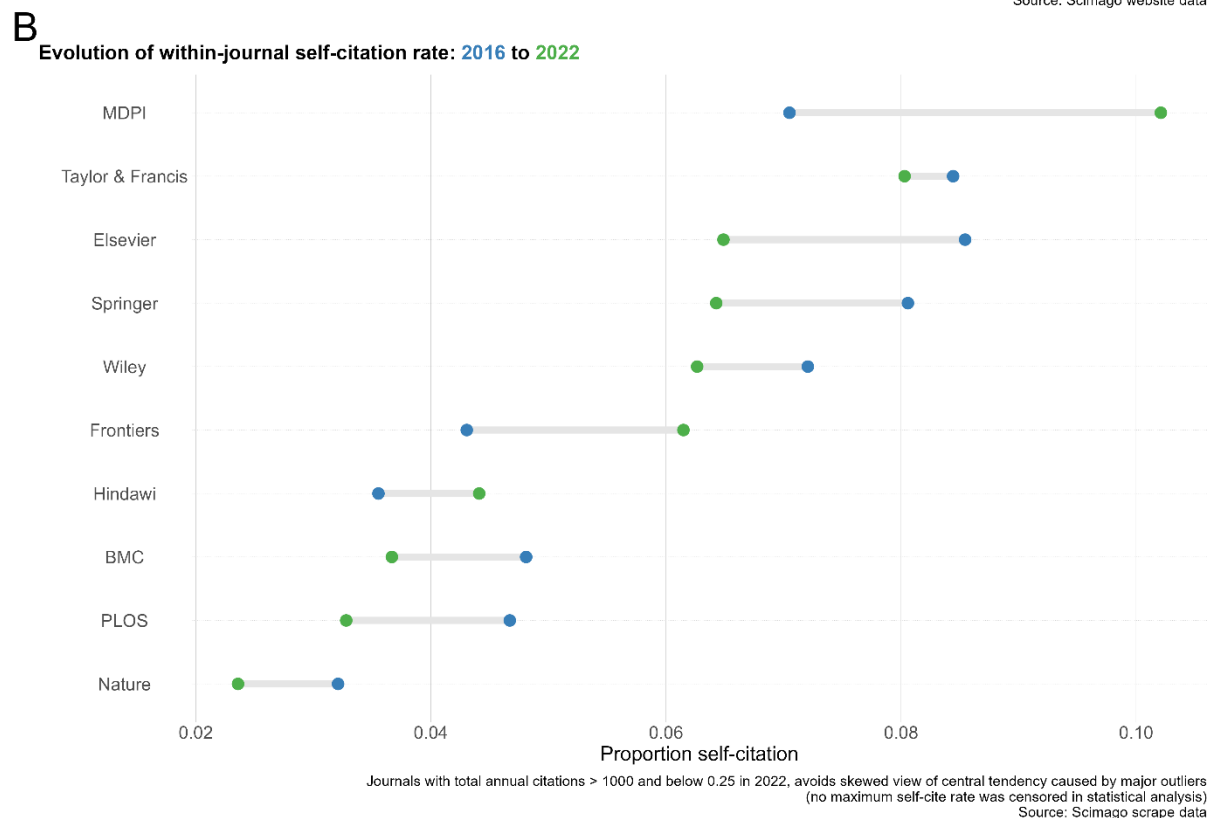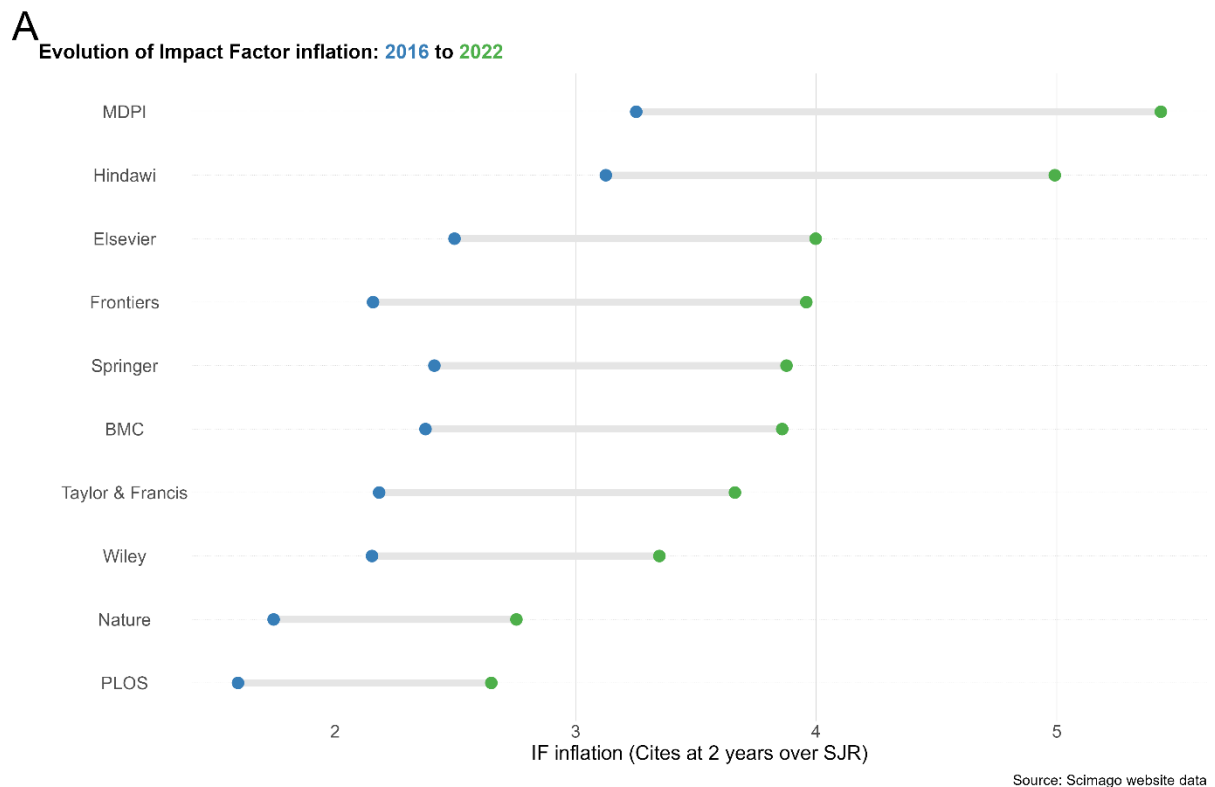
References per document, 2016-22

Fig5supp3: a partial contributor to the increasing total citations being generated is an exponential increase in references per document between 2018-2021. As such, not only is more work being produced (total article growth), but that work is also proportionally generating more citations than articles would be in past years. Here we will note that this growth overlapped the COVID-19 pandemic, which provided an excess of potential writing time to scientists. However, growth in references per doc began already between 2018 and 2019, suggesting the effect of COVID-19 cannot fully explain this change. The ensuing year of 2020 also coincides with acceleration of articles published through special issues (Fig. 2supp1). Thus the growth in references per document is correlated both with a burst of special issue publishing, and world events. References per document also continued to increase between 2021 and 2022 despite measures around COVID-19 relaxing in 2022 – albeit there is indeed a marked decrease in the rate of growth. A full understanding of the influence of COVID-19 on this growth in references per document, and how much references per document explains the universal increase in impact factor (Fig. 5supp1,2) will await data from 2023 where the impact of COVID-19 is further lessened, and normalised for the significant delistings that Clarivate performed in March 2023 that have had a marked effect on the calculation of impact factor (Fig. 5supp4A).
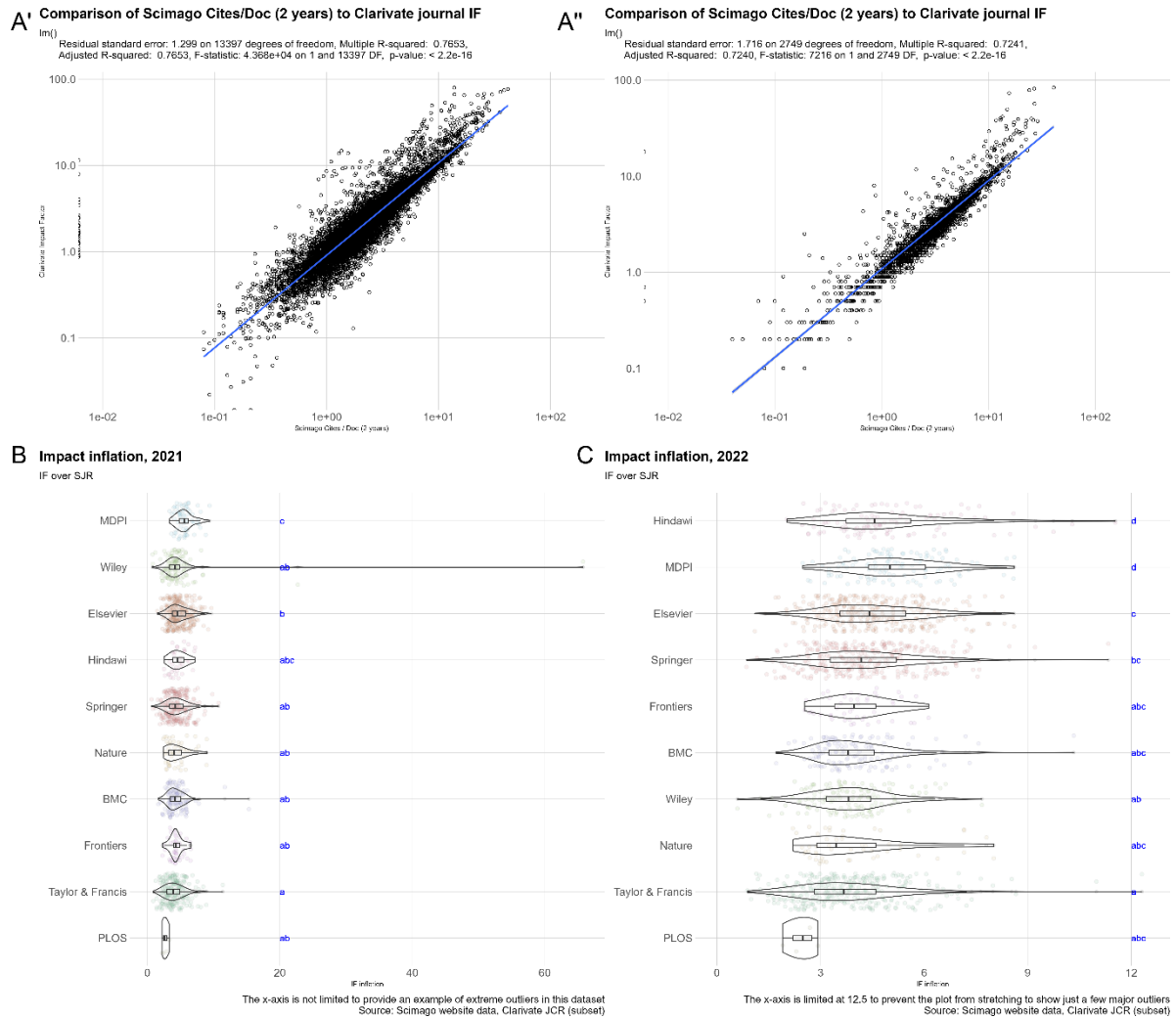
Fig5supp4: validation of Scimago Cites/Doc (2 years) as a proxy of Clarivate journal IF. A) Prior to 2022, "Cites/Doc (2 years)" and Clarivate IF have a correlation of *adj-R²* = 0.77 (A'), but due to mass delistings by Clarivate (but not Scimago) affecting 2022 journal IFs, there was a decoupling of this correlation for 2022 (A'' : *adj-R²* = 0.72). Regardless, Cites/Doc (2 years) informed by the Scopus database is a good proxy of Clarivate Web of Science IF. B-C) Impact Inflation calculated using a subset of Clarivate IFs we could download for our publishers of interest in 2021 (B) and 2022 (C). In both years, MDPI has significantly higher Impact Inflation compared to all other publishers except Hindawi. Here we leave an example in (B) of what is meant by "major outliers" in Fig. 5, to show that plotting the full x-axis range does not change trends, but is aesthetically disguisting.

A

Evolution of Impact Factor inflation: 2016 to 2022

MDPI
Hindawi
Elsevier
Frontiers
Springer
BMC
Taylor & Francis
Wiley
Nature
PLOS

IF inflation (Cites at 2 years over SJR)

Source: Scimago website data

B

Evolution of within-journal self-citation rate: 2016 to 2022

MDPI
Taylor & Francis
Elsevier
Springer
Wiley
Frontiers
Hindawi
BMC
PLOS
Nature

Proportion self-citation

Journals with total annual citations > 1000 and below 0.25 in 2022, avoids skewed view of central tendency caused by major outliers
(no maximum self-cite rate was censored in statistical analysis)
Source: Scimago scrape data

Fig5supp5: evolution of Impact Inflation and within-journal self-citation between 2016 and 2022. A) Impact Inflation has increased universally across publishers (absolute values summarised in Table 2). B) Within-journal self-citation has increased in recent years specifically for publishers that grew through use of the special issue model of publishing: MDPI, Frontiers, and Hindawi. Notably, MDPI has higher self-citation rates than any other publisher, exceeding previous highs from 2016 (Elsevier, Taylor & Francis) by over one percentage point.

Within-journal self citation rate, 2021

Does not include journals citing across each other
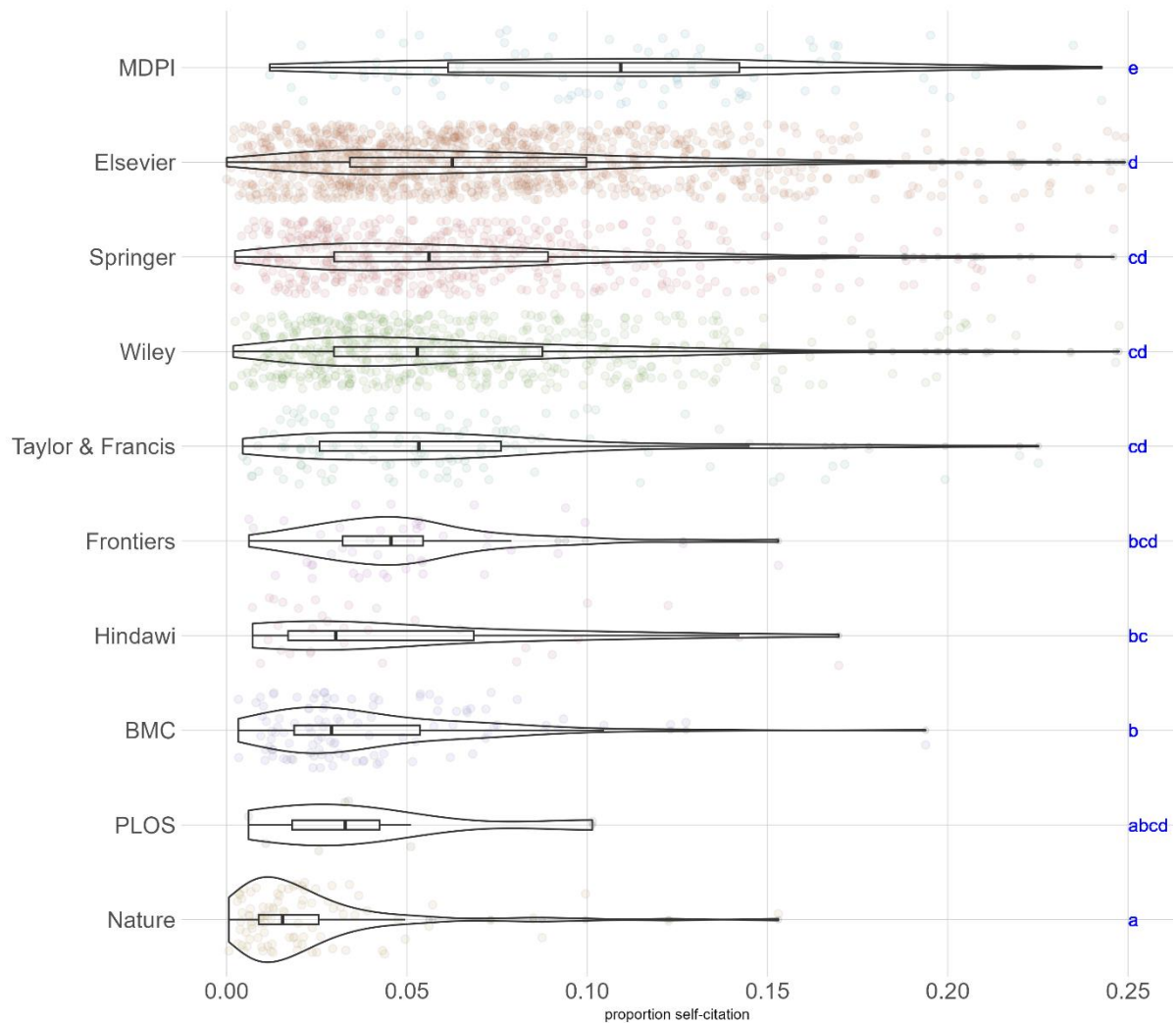
Only shows journals with total annual citations > 1000
the x axis is cut off at 0.25 to prevent the plot from stretching due to a few major outliers
Source: Scimago scrape data

Fig5supp6: within-journal self-citation rates from 2021, supporting the trend in 2022 that MDPI uniquely has significantly higher self-citation rates compared to all other publishers. A difference between 2021 and 2022 is that in 2022, MDPI and Taylor & Francis were not significantly different (*P* > .05). In 2021, this difference was significant (P = 3e-7).
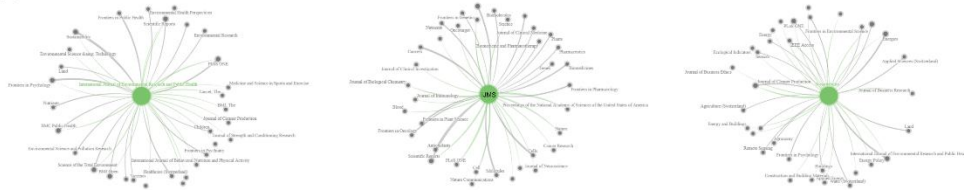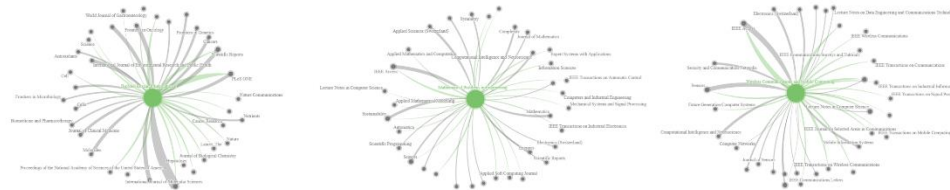
Fig5supp7: example citation networks of single journals from Scimago. Journals were selected from the largest journals by publisher. Journal citation reciprocity depicted with grey arrows for incoming citations, and green arrows for outgoing citations. MDPI journals make up large fractions of the tota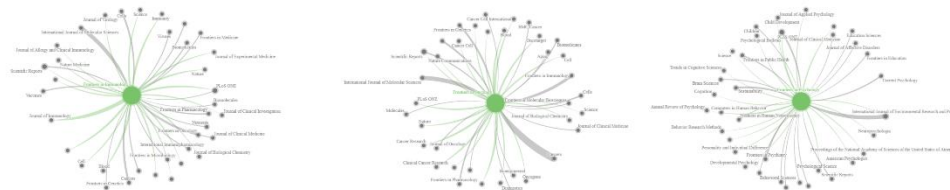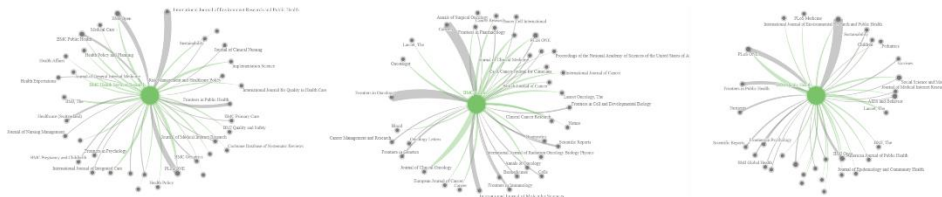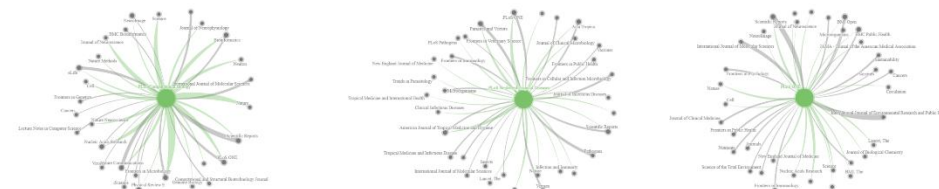l incoming citations of their own journals, uniquely true of MDPI and not other publishers in our analysis. This result is in keeping with MDPI themselves, who reported a ~29% within-MDPI citation rate (shown in supplementary materials and methods). High rates of Impact Inflation of Hindawi journals may come from disproportionate citations received from MDPI journals. For instance, a plurality of citations to BioMed Research International (row 2, column 1) come as large chunks (thick grey arrows) from MDPI journals (International Journal of Molecular Sciences, International Journal of Environmental Research and Public Health, Nutrients, Antioxidants, Cancers, etc…). A similar pattern is seen for Mathematical Problems in Engineering (row 2, column 2): Sustainability, Mathematics, Applied Sciences (Switzerland), Symmetry, Sensors, etc… Because the Scimago Journal Rank metric has an upper limit on the prestige a single source can provide, the large number of citations individual MDPI journals are exporting may be an important factor leading to universal trends in Impact Inflation. *A full-resolution version of this figure is available online at doi: 10.6084/m9.figshare.24203790.*

Table 1supp1: Change in submitted papers relative to the previous month for the 25 largest MDPI journals. On March 23rd 2023 Clarivate announced the delisting of the MDPI flagship journal International Journal of Environmental Research and Public Health (IJERPH), as well as Journal of Risk and Financial Management (JRFM). Following this, submissions to IJERPH plummeted by 73 percentage points in April 2023 compared to March 2023, which already showed a slowdown overall compared to February 2023. Moreover, submissions to MDPI journals in general were down in April 2023 across the board compared to March. A similar pattern was seen in early 2022 following the Chinese Academy of Science release of their "Early Warning Journal List" trial published Dec 31st 2021, which featured multiple MDPI journals. These patterns demonstrate that external authorities, such as Clarivate or national academies of science, can have profound impacts on author submission behaviour, despite opaque methodologies surrounding their decisions to list or delist journals.

Monthly % change in **submitted** papers – 25 largest MDPI journals

Overall = 98 journals with an Impact Factor as of February 2023 and detailed for the 25 largest MDPI journals

JOURNAL	N 2022	2022												2023							
		JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG
Overall	275890	-2.88	-9.71	12.67	-3.20	0.50	0.39	-1.96	7.38	8.50	10.66	5.28	2.38	-0.30	4.29	4.90	-21.13	-3.96	-0.29	-0.61	-1.59
ijerph	17445	-2.62	-8.67	18.49	2.86	-2.42	2.77	-1.93	4.10	12.46	10.59	7.32	1.03	1.20	5.72	-17.76	-73.12	-3.09	-23.21	-8.97	-2.55
sustainability	17394	-4.64	-5.84	17.50	-3.71	2.63	-7.87	-1.13	16.75	2.09	15.11	4.87	-1.50	-2.90	11.13	23.61	-19.57	-2.33	9.50	-5.56	-1.65
ijms	16482	7.45	-15.43	13.13	1.26	0.16	6.34	-7.25	5.49	17.32	12.23	6.42	2.53	-2.69	6.59	-0.54	-18.61	-1.38	2.59	-2.69	3.02
applsci	13229	-8.25	-14.51	4.59	-13.17	9.99	-0.75	-1.35	-6.55	5.67	12.49	8.11	-1.18	8.33	4.20	11.82	-11.55	-2.22	-4.51	-0.53	-4.21
sensors	10149	4.10	-7.35	14.64	-8.35	-1.25	1.65	-2.64	3.26	12.88	7.86	6.65	13.72	-8.77	2.69	5.57	-12.83	-14.23	-1.69	-0.65	-4.12
energies	9843	-11.19	-11.57	9.62	0.00	5.28	-4.32	-2.26	5.20	0.55	28.78	3.57	-0.58	-6.47	5.53	-9.77	-5.94	-11.94	-2.63	4.42	-1.57
materials	9184	-8.12	-14.23	14.64	-11.54	-0.11	1.28	1.27	9.81	9.98	9.68	1.97	4.64	-3.18	-3.89	0.24	-7.36	-9.06	-3.67	1.17	-5.21
molecules	9144	-1.85	-7.43	11.51	-8.76	4.18	-3.38	-1.23	4.11	18.82	10.20	5.19	8.60	2.70	-6.99	2.63	-14.34	-15.79	0.87	1.81	-13.78
jcm	7641	-8.11	-0.10	6.53	-4.34	7.09	11.22	-6.62	5.76	2.01	-0.41	1.73	9.97	22.25	11.27	18.53	-26.60	-11.39	16.23	-4.53	9.94
remotesensing	6479	-1.42	-5.28	11.25	0.09	2.82	-0.71	0.27	14.67	1.24	2.53	5.08	-5.33	-2.48	1.08	9.67	-15.49	1.64	-4.69	3.73	0.65
cancers	6359	1.19	-8.63	7.40	-5.49	1.06	-3.56	7.81	13.58	8.33	7.60	1.06	12.71	-7.74	3.83	-5.01	-22.87	-2.57	1.85	3.93	1.11
polymers	5625	-8.51	-5.09	9.24	-8.63	-4.63	11.84	4.69	11.94	3.26	19.37	3.00	-5.83	-1.98	-1.77	13.38	-23.50	-2.52	-6.09	-0.65	-4.08
nutrients	5405	-6.87	0.00	4.56	3.72	0.99	-4.65	0.90	8.40	3.52	12.13	5.66	0.00	1.15	1.23	5.70	-19.19	6.78	-6.05	13.41	-1.73
mathematics	4931	15.25	-13.24	16.95	14.49	-3.60	-5.76	1.18	9.43	10.94	14.31	-13.66	6.01	0.42	-3.74	14.67	-13.32	1.56	-2.74	-8.17	-4.88
nanomaterials	4540	-5.43	-15.33	15.38	13.14	-4.33	5.43	-0.17	-2.07	8.26	9.09	2.68	-2.46	-14.56	-4.70	-5.66	-8.51	-5.50	-17.00	10.78	-0.73
electronics	4319	-3.51	-8.38	21.07	-6.57	-10.37	-0.78	5.73	10.09	10.70	6.90	19.66	8.39	7.96	10.45	19.02	-20.73	-5.11	-3.42	3.43	-4.25
water	4245	7.63	-11.50	3.02	-3.63	-1.08	4.17	-16.70	19.83	2.61	10.36	3.08	-0.30	-4.34	4.38	3.00	5.53	-12.00	-1.72	1.91	-0.16
foods	4187	-18.32	-13.32	15.37	-10.96	3.79	7.30	12.76	3.47	12.83	7.62	1.56	-7.92	-1.28	11.18	4.33	-15.70	-2.79	2.74	-3.99	-2.64
cells	4181	3.41	-8.91	12.77	-0.75	-10.65	-5.32	-2.92	25.46	7.75	-3.94	21.39	6.90	-0.69	7.88	-6.15	-34.15	9.75	-2.46	-0.97	-2.15
animals	3666	-10.75	-1.90	23.71	-18.47	1.28	-0.84	-6.60	15.72	14.57	4.47	7.57	2.45	-8.51	13.70	-5.02	-8.91	7.63	3.54	-0.89	-0.60
plants	3642	-7.73	-15.51	21.94	-11.07	1.17	6.15	-7.07	7.21	3.64	20.53	4.51	8.50	5.13	-5.01	7.97	-18.69	0.44	-5.39	9.09	-10.03
biomedicines	3311	-9.26	-3.60	12.24	-1.85	0.75	5.98	-24.16	11.40	5.64	5.14	7.14	-6.67	-0.19	17.70	16.16	-16.67	1.82	2.92	11.99	-0.99
diagnostics	3270	-4.26	-11.23	14.32	4.59	-15.97	11.40	-5.76	8.60	1.67	6.56	20.96	6.52	11.04	0.13	10.60	-17.48	5.88	-7.36	-3.75	-7.63
agronomy	3266	14.23	-17.30	7.74	-3.88	-0.81	-4.68	4.70	-6.73	26.91	-0.34	3.81	-1.50	-0.34	-1.87	15.40	-15.59	-1.24	-0.54	-6.33	9.46
pharmaceutics	2906	-4.82	-11.06	11.14	2.10	-3.65	-8.53	10.62	7.03	8.75	13.88	13.60	0.16	2.33	0.76	7.53	-19.33	-10.24	-5.42	-6.13	-0.22

Source: data scraped on the publisher's website